Understanding Others

The Person Model Theory

Albert Newen

According to Interaction Theory (IT), neither Theory Theory (TT) nor Simulation Theory (ST) give an adequate account of how we understand others. Their shared defect, it is claimed, is that both focus on third-person observation of the other, and neglect the role of social interaction. While interaction theory is made to account for the latter, it has problems doing justice to explicit attributions of propositional attitudes, especially from an observational stance. The latter received a new explanation by the Narrative Practice Hypothesis (NPH) which focuses on story-based explanations and tends to underestimate the relevance of nonlinguistic intuitive understanding. In this paper, I first try to do justice to what is plausible about each of the four approaches by accepting that each account introduces one plausible epistemic strategy for understanding others, which leads us to a multiplicity view about the epistemic strategies for understanding others. But it will then be argued that an adequate theory of understanding others needs further adjustment and correction because we need to account for the fact that we usually understand others on the basis of specific background knowledge that becomes more enriched during our life; I thus propose Person Model Theory (PMT) as a fruitful alternative. On my account, understanding turns on developing "person models" of ourselves, of other individuals, and of groups. These person models are the basis on which we register and evaluate persons as having mental as well as physical properties. I argue that person models can be either implicitly represented or explicitly available. This is accounted for by describing two kinds of person model, corresponding to the two ways of understanding others; very early in life we develop implicit *person schemata*, where a person schema is an implicitly-represented unity of sensory-motor abilities and basic mental phenomena related to one human being (or a group of humans); and we also develop *person images*, where a person image is a unity of explicitly-registered mental and physical phenomena related to one human being (or a group). I argue that the person model theory has more explanatory power than the other candidates.

Keywords
Person image | Person model theory | Person models | Person schema | Simulation theory | Theory theory

Author

Albert Newen
albert.newen@rub.de
Ruhr-Universität Bochum
Bochum, Germany

Commentator

Lisa Quadt
lisquadt@students.uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

1 Introduction

A key question for social cognition is: Can we provide an adequate theoretical analysis of the process of understanding other human beings? For over twenty years, there have been only two possible answers to this question—that offered by "Theory Theory", and that of "Simulation Theory". The central claim of TT is that one's understanding of another essentially relies on a folk-psychological *theory*, where some take the position that the relevant folk psychology is innate (e.g., Baron-Cohen 1995), while others claim that it is acquired (Gopnik 1993). In contrast, ST holds that we understand others by means of *simulation* (e.g., Goldman 2006), where simulation can take place at two levels, referred to as low-level and high-level simulation (Goldman 2006). In recent years, however, it has become clear that both positions have significant limitations. One central problem is claimed to be that both TT and ST take a

Newen, A. (2015). Understanding Others - The Person Model Theory.
In T. Metzinger & J. M. Windt (Eds). *Open MIND:* 26(T). Frankfurt am Main: MIND Group. doi: 10.15502/9783958570320

primarily observational stance towards the other when analysing understanding:[1] critics maintain that this observational stance is a nonstandard, intellectual perspective, and that in fact we are normally involved in *interaction* when we try to understand others. Developing this line of thought, Gallagher's interaction theory (2001) combines involvement in interaction with a direct perception thesis, such that we can directly perceive the mental states of others and do not have to infer them. Another alternative proposal is Hutto's *narrative* account of social understanding (2008), on which understanding others relies centrally on telling or understanding stories. These idealized positions are the bases for a wide range of mixed positions, with which I will engage shortly. Yet even if we consider only these idealized positions, a new central defect quickly becomes clear: namely, that these positions offer answers to rather different questions. Thus, in a first step, I aim to reorganize the field of the main positions and use this framework to situate my own view, which I refer to as the *person model theory* (Newen & Schlicht 2009; Newen & Vogeley 2011): this account is characterized by the claim that we understand others by essentially relying on person models of individuals, or of groups.

2 Reshaping the field of positions by distinguishing central questions

The question "How do we understand other human beings?" has to be divided into several subquestions, the first of which is: What epistemic strategy do we adopt to register or assess the other's cognitive states? To reach any kind of assessment of the other we need to obtain information within a concrete situation. The second question is: Once obtained, how is this prior information stored and organized? This second aspect is important, because we always rely on prior background knowledge in our assessments of others. One main defect of the debate thus far has turns on the failure to distinguish these two questions. The debate between the two classic positions, ST and TT, can roughly be described as a misunderstanding stemming from their dealing with different questions: while ST insists that the use of simulation is the standard epistemic strategy, TT insists that the prior information we have about others is organized as a folk-psychological theory. Concerning their main claims, these accounts are not in opposition. The opposition only becomes visible if for each account we consider their favoured answer to both questions. The classic opposition between ST and TT can then be described as follows: TT claims that the epistemic strategy relies upon theory-based inferences, and that the prior information is organized as a folk-psychological theory; while ST claims that the strategy for information-processing involves simulation (to put oneself into the other person's shoes) which draws only on my own experience as the source of data for simulation, leaving it open as to whether these data form a theory.

Before turning to the question of which information-processing strategy we use to understand others, I first provide a brief survey of the field. Thus, in addition to TT and ST, we have Gallagher's IT, which focuses only on the strategy question; it claims that we understand others through social interaction and/or by direct perception, i.e., we can directly perceive mental phenomena; we also have Hutto's account, which is given in terms of story-telling. Their more elaborate joint account combines these claims (Gallagher & Hutto 2008), maintaining that we can distinguish three epistemic strategies for understanding others, depending on the stage of cognitive development in ontogeny: direct perception in very early childhood, followed by interactional understanding, and finally narrative understanding (Hutto 2008). In contrast, my aim will be to show that we actually use a multiplicity of information-processing strategies to understand others, depending on the context; the proposed account, then, is even

1 This is a simplified view. A closer look into Gopnik & Meltzoff (1997) shows that their version of TT accounts for interaction as part of the development of an understanding of action and agency (Chap. 5). But interaction is not accounted for in the further dimensions of understanding others. From a bird's eye view this characterization is not inadequate, although it needs qualification. As the reader will see, my person model theory integrates this initial understanding of action and agency as elements of forming implicit person models that at the beginning may not be rich and abstract enough to warrant being called a theory (see n. 6 below).

richer than the three strategies proposed by the joint account of Hutto and Gallagher.

3 The epistemic strategy for understanding others

3.1 What about simulation?

According to Goldman's (2006) elaborate simulation account, we must distinguish between low-level and high-level mindreading. "Mindreading", in his view, comprises all cases of evaluating the mental state(s) of another person that normally lead to a language-based attribution of a mental state to a person. In the case of high-level mindreading, this is

> [...] mindreading with one or more of the following features: (a) it targets mental states of a relatively complex nature, such as propositional attitudes; (b) some components of the mindreading process are subject to voluntary control; and (c) the process has some degree of accessibility to consciousness. (Goldman 2006, p. 147)

The paradigmatic case of high-level mindreading is understanding another person's decision. Third-person attribution of a decision consists of:

- imagining propositional attitudes in a form of *enactment imagination;*
- using (the same) decision-making mechanisms (as in the first-person case);
- projecting the result of using that mechanism onto a third person by attributing a decision.

We can easily present cases in which these proposed essential steps are not involved. For (i), to understand a person suffering from a delusion of persecution, we are not able to deploy enactment imagination: Their case is just too different from our own experience. And the same may be true in cases of deep cultural difference. For (ii), if I have experience with the other person such that I know that he has idiosyncratic, non-rational decision-making habits when making weekend plans, I can use this knowledge to model his decision and not my own decision-making apparatus, since I have experience that my own apparatus differs from his (at least concerning weekend plans). For (iii), grant for the sake of argument that we have a plausible candidate for the beliefs and desires of the other and we use this for enactment imagination as well as input for my own decision-making apparatus, thus reaching a decision to do action A. Then, according to Goldman, I should project this decision onto the other person. Yet there remains an essential gap, which is noted by Goldman but not adequately addressed by him: He observes the necessity of "quarantining" my idiosyncratic background beliefs if I want to come to an adequate projection of the decision to do action A. Suppose I am warranted in presupposing that the other wants an ice-cream, has money, and that there is a nearby cafeteria where he can get one: then the decision-making apparatus may come to the decision to buy an ice-cream. If, however, I am a person who is extremely parsimonious with money, then my own background desire to save money may prevent me from buying the ice-cream in the same situation, and so this intervenes and I do not attribute the decision to buy an ice-cream to the other. But it seems that the desire to save money is—often, at least—an idiosyncratic desire that I should not use in my projection. Yet how do I know which of my own beliefs and desires are idiosyncratic and do not relate to the person I aim to understand? To solve this problem, I must already possess some view about the attitudes of the other as compared to me; yet this was what we were aiming to understand. In general, then, Goldman's theory of high-level mindreading has difficulties even getting off the ground: It starts by making presuppositions about the beliefs and desires of the other person, where this is exactly what we were aiming to understand. The same problem appears again in the projection phase, as just illustrated. Thus, high-level mindreading is a very special case of simulating a decision of the other, specifically when I already know a lot about the other, which I can use as input. This leaves open the question of how we get this in-

formation at all. Goldman tries to account for problems of this kind by accepting the importance of inference-based strategies and the organization of the prior information in form of a theory. Thus he is no longer developing a pure simulation theory but rather a hybrid account. Nevertheless, the counterexamples are not rare but in fact quite typical, and thus they cast doubt on the typicality and pervasiveness of high-level simulation in mindreading decisions.

Goldman may, however, appeal to his strategy of low-level mindreading, which is characterized as an activity that is "comparatively simple, primitive, automatic, and largely below the level of consciousness" (2006, p. 113). Goldman uses as a paradigmatic case face-based recognition of emotion, and he makes an additional appeal to "mirror neurons", proposing that mirror neurons are not only relevant in the case of understanding motor activities (in both observing and doing them) but also for recognizing mental phenomena like pain and disgust. The most elaborate case relevant to this area of discussion concerns the study of disgust: It has been shown that experiencing disgust and observing disgust are dependent on certain mirror neurons that are activated in both cases (Wicker et al. 2003). Yet what exactly can we learn from this observation? I develop a critical position on the explanatory potential of mirror neurons in two steps. First, I argue that if mirror neurons could provide us with the whole story of how we understand others, this story would not be given as a case of simulation. Second, I cite evidence that mirror neurons do not provide the core part of the story of understanding others in cases of understanding emotions. Let us start with criticism of the claim that low-level mindreading is a case of simulation. Here I mainly rely on lines of criticism worked out by Gallagher (2007), who claims that "simulation is a personal-level concept that cannot be legitimately applied to subpersonal processes" (p. 363). Even if we do not accept Gallagher's claim, the two core features of simulation would be lacking in the case of resonance processes implemented by mirror neurons: There is neither a first-person perspective involved nor a type of pretence that includes a projection from a first-person perspective to a third-person perspective: "Thus, according to ST, simulation involves the instrumental use of a first-person model to form a third-person 'as if' or a 'pretend' mental state. For subpersonal processes, however, both of these characterizations fail" (Gallagher 2007, p. 360). Why are mirror neurons not an essential part of understanding others? They represent a type of action or emotion that is independent from a first- or third-person perspective; but the distinction between self and other is an essential part of understanding others. Thus a simulation process cannot be fully captured in its essential aspects by the mirror-neuron processes (see Vogeley & Newen 2002).

This criticism of high-level and low-level mindreading does not imply that simulation processes never take place: rather, it suggests that it is only so-called high-level simulation that we can characterize as simulation, and also that it is implausible that simulation is the standard strategy for everyday understanding of others. The latter claim is also based on the observation that we often rely on automatic, intuitive understanding of others without any conscious considerations.

3.2 What about theory-based inferences?

The same general line of criticism can be developed with respect to theory-based inferences. Such inferences may sometimes be relevant, but are not always so; neither are they the standard strategy for understanding others. Theory-based inferences are important when we are confronted with cases that we find strange or surprising, i.e., situations where we meet another person suffering from a mental disease which we know nothing about, or where the person belongs to a culture that is radically different from ours. In such scenarios, we consciously build hypotheses about the relevant mental phenomena, as well as about the best behavioural strategy to adopt. But most everyday scenarios in which we understand others are not of this type; quite the contrary, we are generally in-

volved in well-known situations with individuals or types of persons with whom we are familiar. There is an effortless application of our know-how regarding dealing with other humans, without any need to rationalize through theory-based inferences. The reply of the advocate of TT would be: Even if the relevant knowledge-how does not involve an explicit theory-based inference, it is only applicable because we rely on *implicit theory-based inferences*. The criticism of this line of thought is twofold: The status of *implicit inferences* is very unclear, because inferences are defined as relations between propositions; and there is evidence that implicit information processes are often non-propositional in nature. For example, in the case of experts, very often the epistemic strategy in their field is complex visual pattern-matching without any inferences; with their superior organization of knowledge, for instance, a chess expert can rapidly perceive a promising move, or a medical expert can quickly notice an inconsistency in a suggested diagnosis. The process of smoothly using this information mainly relies on fine-grained pattern-discrimination and pattern-matching (Gobet 1997) in the relevant situation, rather than on drawing inferences (which only becomes the case if the expert has to consider problematic situations). This is supported by observations of the way people recall chess positions: When seeing a chess board that contains a real, meaningful arrangement, chess experts excel as compared to novices in recalling positions, but perform no better for scrambled, impossible positions (Gobet & Simon 1996). This indicates that they are able to "see" meaningful patterns that a novice cannot see. They may use this ability in addition to making inferences, but inferences are not so much their basic access strategy as an additional one.[2] If neither the strategy of simulation nor the strategy of theory-based inferences is the standard strategy upon which our smooth, everyday understanding of others is based, what form does epistemic access to others' mental states take?

2 It is important to note that I leave it open whether we have to rely on a package of knowledge we are warranted in calling a theory, since I only discuss the strategy of information processing, not the organization of prior knowledge in experts.

3.3 What about direct perception?

In recent years Gallagher (2008) has argued that our epistemic access to others' mental phenomena is essentially based on direct perception. The mental states of others are not hidden, and need not be inferred on the basis of perceiving others' behaviour; rather, behaviour is an expression of the mental phenomena that, in seeing the behaviour, is also seen directly. What does the claim of direct perception involve? Gallagher explains his main idea with an analogy: I can directly see my car. It would be inadequate to claim that I only directly see the colour, the shape, and the material, and then have to infer that it is my car. This is also supported by the fact that, when seeing the car, I at the same time see its drivability. This view does not deny that object-perception involves complex and partially hierarchically-organized brain processes, but it introduces the notion of "smart" perception: If I have learned the concept CAR and I am used to driving cars, I can see a car directly; and in seeing my car I may also see concomitant affordances such as its drivability. The same is true in the case of understanding others: according to Gallagher, by seeing their face and body posture in a specific situation, I can directly see that someone fears an aggressive dog. This can be realized by visual pattern-matching without inferences (see footnote 3 and Newen et al. forthcoming). This is a convincing comparison, especially as regards its potential to give a unified account of both basic perception and what Gallagher calls "smart" perception. The latter are cases in which it appears plausible to accept that perception can be modulated by conceptual information, these usually being described as cases of cognitive penetration (see Macpherson 2012; Vetter & Newen 2014).

Let us illustrate both the basic and the smart perception of an emotion. Basic perception of an emotion takes place when we see fear, joy, anger, or sadness in the face of a person while relying mainly on a single feature, or

small group of features, connected with facial expression (Ekman et al. 1972).[3] This can be done through a bottom-up perceptual process that involves almost no top-down influences, especially if the facial expression is very characteristic of an emotion pattern. In the case of smart perception, the perception of the emotion is modulated by higher-order cognitive processes. To show this, we need a case in which the same facial input leads to a different perception of an emotion as a result of conceptual input. Such cases have indeed been discovered: If we first hear a story describing a very unjust situation that makes us expect the person we are going to see to be angry, we have a strong tendency to see a typical "Ekman" fearful face as an angry face: for example, if I am told that the relevant person made a reservation at the restaurant, waited for an hour while many other people who had come in later were served first, and that after a further hour was informed that she would have to wait for at least another hour, then I have a strong expectation of seeing anger. This has been shown to make us see a typical fearful face as an angry face (Carroll & Russell 1996). *Smart* perception of an emotion is a cognitively-penetrated perception of an emotion, and it is also important for seeing more complex emotions that do not have the typical Ekman facial expressions: if I know that John is jealous of Peter, because he told me so, and I have seen several episodes of Peter behaving intimately towards John's wife Anne, and the next day I see another episode of John flirting with Anne while Peter observes them, I can directly see the jealousy in Peter's face. There is no need for inference-based evaluations. This is parallel to Gallagher's case of seeing one's car: we may describe both cases as cases *of seeing as*: seeing my car as a car (by knowing which affordances come with it) and seeing John's face as evincing jealousy. I illustrated these cases of direct perception because I think Gallagher makes an important point when he claims that the main source of understanding others is direct perception (whether basic or smart). Nevertheless, there are clear limits to direct perception as a form of epistemic access.

Although Gallagher has in the past shown a tendency to overgeneralize the importance of the role of direct perception (2008), he is well aware that there remain cases that cannot be accounted for without going beyond direct perception. This is the case especially concerning our understanding of propositional attitudes—e.g., someone's desire to take a summer holiday with his elder brother in western Turkey. Propositional attitudes are normally radically underdetermined by expressive elements such as facial expressions, gestures, body postures, etc., in a given situation. In general, therefore, complex human cognitive phenomena of this underdetermined type are communicated by linguistic exchange, or else have to be inferred or simply guessed on the basis of available information. The latter often happens in situations of non-transparent communication due to norms in social situations, or due to the fact that at least one person wants to hide her beliefs and intentions. Since these situations are also part of our everyday life, inferential processes remain part of our everyday understanding of others. Thus, although direct perception is a very important epistemic strategy that we may use in cases of face-based perception of emotion, even "smart" direct perception is not the basic strategy employed to understand complex beliefs, desires, and intentions of others. The latter require inferential processes as well. Thus, we are left with three strategies (simulation, theory-based inferences, direct perception), where none is a clearly dominant standard strategy relevant to all mental phenomena.

But there is at least one further candidate we should take into account, namely *understanding though primary interaction* (Gallagher & Hutto 2008). All the epistemic strategies dis-

3 Although our basic perception mainly relies on certain central cues—e.g., wide-open eyes for fear—the fearful face is not recognized only in one central feature of the face. It requires the integration of several facial features, and not static ones alone. The perceiver also benefits from noticing dynamic visual features like gaze direction: If the gaze is directed away from the perceiver instead of towards her, then this makes the recognition of fear occur faster (see Adams & Kleck 2003; Sander et al. 2006). Together with colleagues I have argued elsewhere that emotion recognition is essentially a process of pattern recognition (Newen et al. forthcoming). This is true for these basic perceptions of emotions. The face is integrated with body posture, since facial expressions are categorized as expressing a specific emotion most rapidly when they are paired with emotionally congruent body postures (Meeren et al. 2005; van den Stock et al. 2007).

cussed so far can apply to situations in which I am simply observing the other without being involved in any interaction. As we have already mentioned, Gallagher views this as a radical defect of such accounts; intuitive understanding of others is part of our everyday life, and this is especially the case if I am not in a purely observational situation but am directly involved in some kind of interaction. Intuitive understanding may then be characterized just by the fact that I notice a social act being directed towards me and so start to interact, such that a standard social interaction is realized, which may be non-linguistic but may also involve linguistic communication—e.g., friendly greetings exchanged while arranging ourselves in line at the office coffee machine. Such a strategy of understanding can only be dominant if the interaction is situated within many conventions, such that smooth understanding can take place without theoretical considerations about the others' beliefs and intentions (de Bruin et al. 2012). But is understanding though primary interaction, as it already takes place in neonate imitation (Meltzoff & Moore 1977, 1994), really the main or the standard strategy for understanding others? Again, even if we grant that this is an important strategy in basic understanding of others, even in adults—e.g., in minimal understanding deployed by smoothly interacting with a stranger who is taking the same bus—we need more advanced strategies to frame estimations about the ramifications of the situation—e.g., whether taking this bus in an unknown city, by night, and with such people on board, is a reasonable risk to take.

3.4 The multiplicity view

To summarize thus far. We use at least four epistemic strategies to understand others, and we learn to use these strategies on the basis of evidence of successful application in the past in relevantly similar situations. We prefer to use simulation strategies where we have evidence that the other is similar to us in respect of many features that are relevant to the situation of evaluation. We typically use theory-based inferences if we need to account for complex mental phenomena or if an intuitive understanding is, for whatever reason, not available. We use understanding by primary interaction in cases in which we are involved in interaction with the other and only need to understand her or him to a limited degree, such that acting according to conventions is sufficient for a smooth interaction. Finally, we normally rely on direct perception of mental phenomena when we are in an observational stance towards the other and have a rich, well-organized body of experience that allows us to recognize mental phenomena as patterns. This is rather easy in cases of emotion recognition, more complex in recognizing intentions, and almost impossible in understanding complex propositional attitudes of others. Only the combination of all four strategies, in full sensitivity to the context and applied on the basis of our experience in successfully using the strategies, makes us experts in understanding others. Thus, we have reached a first main conclusion concerning strategies of understanding, this being what I call the multiplicity view:

> The multiplicity view $=_{\mathrm{Df}}$ There is no standard default strategy of understanding others, but in everyday cases of understanding others we rely on a multiplicity of strategies that we vary depending on the context and on our prior experiences (and which are eventually also triggered by explicit training).[4]

This thesis is also supported by a closer look at mental disorder such as Asperger's syndrome, which is a variant of autism (Fiebich & Coltheart under review). People with Asperger's syndrome lack an intuitive understanding of others. They are unable to directly perceive emotions on the basis of facial expressions, and they tend to avoid social interaction (Vogeley 2012). Thus intuitive understanding by primary interaction or direct perception is not available for them. Since they also tend to experience themselves as being different (Vogeley 2012), they do not use simulation as a strategy: so

4 This view was worked out in parallel by Anika Fiebich in her PhD thesis, under my supervision. She applied the thesis in discussing the case of autism (defended January 2013).

they are left principally with theory-based inferences (Kuzmanovic et al. 2011). And this is what we can observe: persons who are autistic try to understand others by asking for theoretical guidance; thus they might ask how long one is allowed to look into the eyes of another person (Kai Vogeley, personal communication; his expertise is based on regular treatment of more than 300 patients). They also learn what people think in typical situations, but become lost in new situations. Since we have to deal with new situations almost every day, autistic people notice their tendency to get lost and many of them avoid social encounters. This special situation is explained by the fact that in contrast to the usual multiplicity of strategies of understanding, they are left with theory-based inferences alone. People with Down's syndrome are in a contrary kind of situation: they have a good intuitive understanding of others' emotions, but, due to typically very constrained cognitive abilities, they lack any theory-based inferences. In the early years of childhood—where cognitive skills are not so important as in kindergarten or school—their social life is very similar to the social life of children without Down's syndrome; but in later life the interdependence of social interaction with cognitive abilities leads to more problems in building an inclusive social life (Buckley et al. 2002). Thus, the normal multiplicity of strategies may be strongly constrained in some conditions of mental disorders. Furthermore, we can roughly cluster direct perception and interaction as the main epistemic access for *an intuitive understanding* of others, while *inference-based understanding* is based mainly either on a (high-level) simulation strategy or theory-based inferences (including inferences from narratives, see below). Since in our everyday life most of what is going on is intuitive understanding of others, it is especially important to highlight the relevance of social perception. In what follows, I will argue that the most important unit of clustering information about others is neither a facial unit nor an emotion type (or some other subpersonal unit), but the *whole person*—and thus a primary aspect of epistemic access is our ability to perceive persons. We *perceive persons* and their mental settings mainly by directly perceiving them, and/or interacting with them. In addition, we can come to *judgments regarding persons* by simulating them and/or through inference-based understanding.

4 The organization of relevant background knowledge about others

We can now address the second independent question concerning understanding others: How do we organize the information about other people that we already have? This question presupposes that in standard cases of understanding others we are not in a situation in which we are bereft of relevant background knowledge. Quite the contrary: most of the time, we interact with people about whom we have a lot of background knowledge—family members, colleagues, friends, etc. Furthermore, we have background knowledge about the general needs of human beings, the special needs of students, homeless people, etc. It seems clear that we are relying on this type of knowledge in an essential way when we understand others. There may be very short period as a newborn baby when we start from scratch, armed only with certain inborn minimal mechanisms such as neonate imitation. Even the social smile developed with two months is dependent on external stimulation and learning processes, and babies very quickly start to react selectively towards familiar and foreign individuals. They also expect a typical behavioural interactive pattern from the caregiver. If a mother stops reacting intuitively through normal facial expressions and gestures, and instead reacts with a "still face", then the baby quickly starts to cry (Bertin & Striano 2006; Nagy 2008). The baby is irritated by the unexpected pattern of reaction. How, then, are all these different types of background information about the other organized and used in social understanding?

4.1 Are we organizing our prior knowledge in folk-psychological theories?

The question of whether we are organizing our knowledge according to folk-psychological theor-

ies has received a number of different answers. According to TT, this is exactly what happens. In understanding others we rely on folk-psychological rules such as: "If she desires an ice-cream and she believes that she can get one with her money at the cafeteria, then she will go to the cafeteria". No doubt folk-psychological rules, organized according to a belief–desire psychology, are an important instrument for understanding others; but they are by no means the only one. Often it is sufficient to know the conventions in a society to understand what someone is doing and will do next, e.g., if someone is in Japan and he enters a restaurant, he will first take off his shoes, then take a seat, and then will be asked to order. So, seeing someone entering a restaurant who looks like a guest (and not a waiter) allows us to expect a specific conventionally-regulated sequence of behaviour. If one has a liberal notion of folk-psychological theory, then we may add such behavioural conventions into that theory. But even then the question remains whether our understanding of others always relies on knowledge organized as a folk-psychological theory. A counterexample can be proposed by reference to cases of basic intuitive understanding: e.g., the still-face reaction by the caregiver, instead of a typical smiling facial expression and gestural response, makes the baby start to cry (as we saw above). There is thus an intuitive recognition of basic emotions like fear, anger, happiness, or sadness. This may rely on inborn emotion recognition mechanisms, or mechanisms learned very early, which may be evolutionarily anchored, since recognizing such basic emotions is essential for survival (Griffiths 1997; Panksepp 2005). There are two ways in which the counterexample might be blocked: (i) It could be maintained that some folk-psychological theories are inborn (Baron-Cohen 1995) and that intuitive understanding such as face-based recognition of emotion already involves a theoretical package. The problem with this line of reasoning is that the notion of theory, stretched that far, starts to look very implausible. A theory is constituted by a minimal package of systematically interconnected beliefs; and even if a belief is understood in a liberal way such that it does not presuppose linguistic representations, it remains highly questionable whether basic cases of faced-based recognition can be characterized as a systematically interconnected set of beliefs. The standard descriptions of face-based recognition of emotion (e.g., Goldman 2006) on a neural level highlight the relevance of mirror neuron mechanisms and characterize the underlying mechanism as a rather basic and partially independent pattern-recognition process, and thus as not forming a theory. A defect in recognizing disgust does not automatically lead to a defect in recognizing other basic emotions like happiness or sadness (Calder et al. 2000). (ii) A more promising move is to claim that the folk-psychological theory is learned (Gopnik 1993). This view is compatible with some basic processes of understanding which do not yet form a theory, but are developed into one as they are integrated step by step into a systematically-organized body of knowledge. This is a plausible and to some extent empirically grounded view (Gopnik & Meltzoff 1997; Newen & Vogeley 2003).[5] One shortcoming of this view, however, is that its proponents tend to appeal to examples that have a strong focus on general folk-psychological rules, such as: "All humans need to drink, thus if someone picks up a glass in the kitchen, he intends to pour into it some liquid to drink". This neglects a very important phenomenon, namely that we mostly interact not with complete strangers but with persons we know at least partly and often very well. For example, if Michael observes his son in the kitchen grasping a glass he does not appeal to the folk-psychological rule at all, since he knows that his son—despite his education—still only drinks from a bottle when at home, and that if he takes up a

5 Gopnik and Meltzoff insist that the basic registration of objects—e.g., their being sensitive to object permanence, as well as the basic registration of agents rooted in their being able to distinguish inanimate objects and living beings—which babies develop very early on, shows that they already have an *initial theory* of objects and agents. They argue that the already innate "structures are rich enough and abstract enough to merit the name of theories themselves" (Gopnik & Meltzoff 1997, p. 82). But it is questionable whether the notion of theory really has any fruitful role here, because, for example, explanations and predictions of the behaviour of a baby when seeing an object are extremely constrained. The developmental story told by Gopnik and Meltzoff is of course very plausible and at some point turns into a theory, because the transformation of the representation in the context of new cognitive abilities comes with a rich and systematic package of explanations and predictions.

glass it is just because he wants to use it for practising magic tricks. This indicates that all the theories canvassed thus far have a blind spot: so far it seems simply to have been neglected that we rely extensively on knowledge of properties of individuals, which is organized as belonging to one specific individual (the son, the partner etc.) or to a group (students, managers, etc.). The general worry concerning the organization of this knowledge, according to TT, can also be expressed as follows: How are we able to apply a general theory of typically human features in a *specific social* situation? If we want to integrate our prior background knowledge of persons as individuals or as belonging to a group, e.g., to a profession, then we can characterize the organization of this knowledge as *person models*. Person models of individuals and groups are by far the most important source of understanding others, I will argue, and since they involve specific knowledge, they are the natural candidate for enabling adequate deployment of more general knowledge of human psychology in concrete everyday situations. It remains to be discussed, then, whether person models have the status of a folk-psychological theory or not. To adumbrate my line of argument: no doubt some elaborate person models are systematically-interconnected sets of beliefs, but not all of them have to be, because some person models only involve very sparse and basic properties that are not highly interconnected.

4.2 Do we organize our prior knowledge in narratives?

As we saw earlier, one recent account of understanding others, proposed by Dan Hutto (2008), holds that understanding others mainly relies on telling stories and using this knowledge to understand individuals. The core claim of his NPH (Narrative Practice Hypothesis) is

> [...] that direct encounters with stories about persons who act for reasons—those supplied in interactive contexts by responsive caregivers—is the normal route through which children become familiar with both (1) the basic structure of folk psychology and (2) the norm-governed possibilities for wielding it in practice, thus learning both how and when to use it. (Hutto 2008, preface, p. x)

One focus of his theory is not so much how the prior background knowledge of others is organized, but rather how children are able to acquire it. His developmental claim is that the central route for learning relevant background knowledge is listening to stories about persons. I grant that this is an important additional route of epistemic access to relevant knowledge about others; but it is already an advanced method, not normally used before the second year of life. Furthermore, in such cases the focus is not epistemic access to knowledge used to understand the other in the situation (i.e., when listening to the storyteller), but rather to gain new background knowledge with an eye to future understanding of others. In a follow-up paper written together with Gallagher (Gallagher & Hutto 2008), Hutto and Gallagher enrich their views about epistemic access through appeal to direct perception and interaction (see above) in addition to learning by narratives. It is important to note the difference between epistemic access to information that allows me to understand the other in the actual situation (see section 3) and epistemic access to background knowledge relevant for future usage. Thus, by granting that narratives are an additional instrument for learning about important properties of persons, I can enrich my multiplicity claim as characterized above. In integrating this idea, one should also generalize it: we not only learn important background information that helps us to understand others by listening to stories told by a caregiver, but also by reading stories, especially novels.

Let us now briefly discuss the NPH considered as a claim about the organization of our background knowledge. If I have elaborate and explicit knowledge of a person, I may have acquired it by listening to or reading a story, and I may tell a story if someone asks me about this person. But, as the interaction view highlights, sometimes my knowledge may be anchored in

the interaction, yet still be non-linguistically represented, and only activated in similar interactive situations. Our rich non-linguistic knowledge about other human beings, which we acquire when directly perceiving them (tone of voice, what they look like) or interacting with them, or when realizing a joint action, etc., are often not linguistically coded and thus not memorized as a linguistic story. If we widen the notion of a story such that it includes any sequence of memorized events, we lose track of any interesting notion of "story". In fact, we are instead going in the direction that I propose, i.e., that we organize our prior knowledge about others through unifying it in person models. Some such models may include properties of a person that are connected as or with stories, but the core of a person model is a unity of features of a person that are grouped together as belonging to one individual or to a group, where the features may be as primitive as the tone of voice of a person, and have no connection to any story, even in a wide sense.

Although our prior knowledge about others is the main component of our understanding of others in a specific situation, most of the theories canvassed above did not present any clear view on how this knowledge is organized.[6] We found only two suggestions: relevant prior knowledge is organized either as a folk-psychological theory or as a narrative. Neither proposal covers all relevant cases: neither accounts for the innate or very-early-learned (nontheoretical) basic background knowledge that enables us to effect smooth interaction and allows us to rely on a basic intuitive understanding of others. And, furthermore, as I argue in the following, there is an alternative view, the person model theory, which is able to integrate the plausible aspects of these two suggestions, and additionally allows us to explain a variety of phenomena that the alternative views did not or cannot take into account—especially the integration of features of other human beings that allow us to realize an intuitive understanding of them.

5 The person model theory

Before expounding the new account, let me highlight two main criteria of adequacy for any plausible candidate theory and some open questions. (i) The theory should account for two levels of understanding others from a phenomenological perspective, namely intuitive understanding and inference-based understanding. This was first clearly discussed by Gallagher (2001), while Goldman (2006) described it in his distinction between low-level and high-level mindreading. What, we may then ask, would be an adequate way of establishing this distinction? (ii) We learned from Gallagher (2005) that we should distinguish understanding others by observation from understanding by interaction.

There are also a number of open research questions that can potentially be answered in developing the alternative account: (a) What is the relation between understanding oneself and understanding others? Here the ST claims that understanding oneself is the basis for all understanding of others, while TT is neutral; Carruthers, for example, has famously argued that understanding others is the source of our self-understanding (2009). (b) What is the relation between understanding persons and understanding objects or situations? (c) How can we best account for the difference between understanding a well-known person, on the one hand, and a complete stranger, on the other?

The new alternative theory, which promises to deal with these open questions, is the person model theory. The central claim of this theory is that we organize our prior knowledge that is used to understand others into something we can call person models, and that accounting for our way of using person models is the most informative factor when analyzing our everyday understanding of others. A person model[7] is a unity of properties or features that

6 This includes, e.g., the ST, which mainly offers a claim about how we use our knowledge to understand others, and that the main source of this knowledge—in addition to situational input—is one's own experience. But a representative of ST can easily grant that relevant prior knowledge is organized in a folk-psychological theory. She only insists that the strategy of application of this knowledge in a situation is a simulation process.

7 An important question which I cannot discuss in this paper is the question of the development of person model and the limits of application. Some very sketchy remarks may be of help here for urgent

Table 1: Varieties of person models

Person models	Self	Other: Individuals	Other: Groups
Person schema	Self schema	Individual person schema	Group person schema
Person image	Self image	Individual person image	Group person image

we represent in memory as belonging to one person or a group (resp. type) of persons. To account for the difference between two types of understanding others (intuitive versus inference-based understanding), I suggest that there are two types of person models in use: implicit person models,[8] which we shall call person schemata; and explicit person models, which we shall call person images. Very early in life we develop *person schemata*: a person schema is an implicit person model and can typically be described as a unity of sensory-motor abilities and basic mental phenomena[9] realized by basic representations and associated with one human being (or a group of humans), where the schema typically functions without any explicit considerations and is activated when directly seeing or interacting with another person. A person schema is thus the unity of implicitly-available information about a person that is thus not easily accessible in terms of being reportable but is nevertheless used in a specific situation. In other words, a person schema is the basic unit that enables a practical knowledge (a *knowledge how)* for dealing with another human being while this ability relies mainly upon social perception and interaction. Person schemata can be developed step by step into *person images*. A person image is a unity of explicitly represented and typically consciously available mental and physical phenomena related to a human being (or a group of people). Thus, a person image is the unity of rather easily and explicitly available information about a person, including the person's mental setting. Both person schemata and person images can be developed for an individual, e.g., one's mother, brother, best friend, etc., as well as for groups of people, e.g., medical doctors, homeless people, managers, etc. Furthermore, person models are created for other people but also for oneself.[10] In the case of modelling oneself we can speak of a self-model that we develop implicitly as a self-schema and explicitly as a self-image. Thus, we have the following varieties of person models (see Table 1).

Person models are characterized here as memorized units of person features, ignoring the difference between long-term or short-term memorization.[11] Person models are distinguished

questions: Concerning the development I suggest that person model enfolds gradually from an early model of living agents which is based on sensitivity for clusters of features indicating animacy and agency. This "agent models" enfold into person models which are systematically enriched by the features I describe as belonging to person schemata and person images. Furthermore, a creation of a person model (which is a unity of information clustered together) does not presuppose a concept of a person. Person models are developed in fact if some typical features of adult healthy human beings are clustered to model an individual or a group of entities which are relevantly similar to adult healthy human beings. Typical core features are e.g., 1. being an agent, 2. being a sentient being, 3. having some minimal control of action. We use person models to understand babies and pets since we usually perceive them as having a minimal amount of core features.

8 I am only presupposing a minimal consensus on using the distinction of implicit versus explicit. It indicates a (gradual) difference in epistemic access such that paradigmatic cases of explicit contents are easily accessible (by the subject's experience, memory, thinking, imagining etc.) while paradigmatic cases of implicit contents are very difficult to access by the subject while they nevertheless influence the subject's cognition and behaviour. Intuitively, explicit content are correlated with our intuitive understanding of *conscious accessibility*, but since the latter is scientifically pretty unclear, I do not want to ground the implicit/explicit distinction on the difference between being or not being consciously accessible.

9 Mental phenomena have different ontological types: states, events, processes, and dispositions. So not only are stable mental phenomena included but so are situational experiences (like tokens of perceptions, emotions, attitudes, etc.).

10 The distinction between *person schema* and *person image* is based on Shaun Gallagher's distinction between *body schema* and *body image*. Establishing a *person schema* of my own body amounts to Gallagher's *body schema*, while a *person image* of my own body is similar to what he introduces as *body image* (2005, p. 24).

11 In a more detailed explication of the theory, it would indeed be useful to distinguish short-term person models (only stored in working memory) and long-term person models (stored in a long-term memory). In addition, other established distinctions in memory can be used to characterize the content of person models, such as procedural and declarative contents as well as episodic and semantic contents. I will, however, ignore these distinctions in this paper.

from the result of understanding in a situation, which may be either a person impression that mainly relies on person schemata, or a person judgment that mainly relies on person images. Let me illustrate one clear virtue of adopting the distinction between person schema and person image by reference to the fact that it can account for the difference between intuitive understanding and inference-based understanding of others.

5.1 Person schemata

In detail, then, what are person schemata? A *person schema* is an intuitively formed, implicit model of a person; it is a memorized unity of characteristic features of a person including facial features and expression, voice, moving pattern, body posture, gestures, and other perceivable features of a person. The function of clustering these features is to allow us to evaluate a person very quickly in a situation according to evolutionarily-important aspects: is a person familiar, dangerous, aggressive, helpful, or attractive? The evaluation is either expressed in a type of interaction, or it can simply be memorized in an implicit unitary structure for future retrieval, including recognizing the person and activating the former evaluation (Reddy 2008). Our main access to others in everyday life is through perceiving a person and forming an impression (see the review published as a book chapter by Macrae & Quadflieg 2010). To form a person impression, (i) we typically pick up these basic features by means of a quick visual evaluation, even when seeing a person for the first time, where (ii) most features are directly associated with socially-relevant information, and (iii) they are clustered at the level of perceiving the whole person. Let me offer some support for all three characteristics of the process of forming a person impression in a situation that is memorized as a person schema:

(i) *Quick evaluation even with parsimonious information*: Evaluations of threat (which is of strong evolutionary relevance) can be made on the basis of exposure to an unfamiliar face lasting as little as 39 milliseconds (Bar et al. 2006). If the exposure to the unfamiliar face lasts about 100 milliseconds, we are able to evaluate likeability, trustworthiness, competence, and aggressiveness with subjective reliability levels that are similar to those generated under longer viewing times (Willis & Todorov 2006).[12]

(ii) *Most features are associated with socially relevant information*: looking into the face is a very rich source of information about a person. Between 3 and 7 months of age, infants learn to recognize the face of the mother and to distinguish it from the faces of strangers, and they start to categorize people according to emotional expression and sex (Nelson 2001). One important source of information that children use from 4 months onwards is the gaze-direction of a person, it having been shown that they can distinguish a direct from an averted gaze (Vecera & Johnson 1995). From 9 months onwards, infants learn to register the joint attention of the infant and an adult as directed towards an object (Cleveland & Striano 2007). Thus, on the basis of gaze-interaction they evaluate whether joint attention towards an object has been established or not, and learn to direct the attention of the other if necessary (Tomasello 1999). Between the ages of 9 and 18 months, children start to use gaze-information to register the *goal* of the action of the other human: they attend immediately to the eyes when the intentions of an actor are ambiguous (Phillips et al. 1992).

Let me now pick out some results based on studies of adults that illustrate the informational value of single cues. To start with facial expression: in emotion recognition, highly in-

12 The time course can be observed in ERP studies. These studies all support claims about the early information processing of faces, although there is an ongoing debate about how best to interpret the results. The main observations are enhanced responsiveness to faces relative to a variety of other objects with peaks at approximately 100 milliseconds (Herrmann et al. 2005; Liu et al. 2002; Pegna et al. 2004), 170 milliseconds (Bentin et al. 1996; Eimer & McCarthy 1999; Itier & Taylor 2004), and 250 milliseconds (Bentin & Deouell 2000; Schweinberger et al. 2004) after stimulus onset. (For review see Macrae & Quadflieg 2010). Whole bodies (without faces) are evaluated with a delay of 20 milliseconds compared to the evaluation of faces (Gliga & Dehaene-Lambertz 2005). Concerning faces with emotional expressions, the following rather stable result is reported: there is a frontocentral positivity as early as 120 milliseconds after stimulus onset and a later more broadly distributed positivity beyond 250 milliseconds; both are modulated by emotional facial expressions (Eimer & Holmes 2002; Holmes et al. 2003; Vuilleumier & Pourtois 2007; Williams et al. 2006).

formative features include knitted eyebrows for sadness, a smile for happiness, and a frown for anger (Ekman 1972, 1999). To prevent this remark giving the wrong impression, I here highlight some individual features and will argue in the next step that they are part of an integrated view at the level of persons. Salient biological visual markers allow us to easily identify the "big three" categories in person perception (Brewer 1988; Fiske & Neuberg 1990), i.e., sex, race, and age. In the same way, we can illustrate highly informative single features such as body posture: if the other is bending her head in a communicative context, this is unconsciously registered as signalling sympathy (Frey 1999).[13] One important data source here is biological motion-detection as investigated by point light studies. If a person has lights on her hands, feet, and ankles, and some other significant parts of her body, we can videotape her bodily movement in the dark. Such artificial pure biological movement information allows us to register social features, e.g., we can recognize emotions (Ambady & Rosenthal 1992) and attribute personality features (Heberlein et al. 2004) on the basis of seeing dynamic movements alone. Furthermore, there is evidence that social information can be taken from the combination of gesture and body posture alone. In an intercultural study (Bente et al. 2010), an interaction between an employer and an employee (played by two students of one type of culture) was filmed for a short period. Then the film was edited to show only gesture and body posture. This was realized by showing idealized wooden puppets, representing the real interaction while abstracting from facial information, speech, clothing etc. The question to be addressed was, what we can read from seeing the body postures and gestures. The interactions were filmed with students from UAE (United Arabic Emirates), Germany, and the United States; and the test subjects were also drawn from all three countries. With this film, people could determine whether the people in the scene were nervous or not, as well as the dominance relation, i.e., they could see who was the boss. This is an interculturally-shared social understanding of otherwise culturally variable cues of body posture and gesture (the US students moved a lot while the UAE students moved rarely). They furthermore could perceive the level of friendliness in the interaction, although the study showed that we are good at this only in assessing our own culture.[14] Furthermore, there are many more complex culturally-dependent visual features that (according to other studies) we use for evaluating the other—e.g., physical attractiveness, where attractive people are evaluated as possessing more desirable characteristics than their less attractive counterparts, a phenomenon that has been labelled the *beauty-is-good stereotype* (Dion et al. 1972; Eagly et al. 1991). These kinds of stereotypes are especially connected with racial classifications: African–Americans are stereotypically assumed to be lazy, criminal, and uneducated, but also musical and athletic (Devine & Elliot 1995), whereas Asian–Americans are considered to be intelligent, industrious, conservative, and shy (Lin et al. 2005). Most observers in our culture assume that people with stylish hair and extravagant clothing are highly extrovert (Borkenau & Liebler 1992). We live with a lot of these deeply culturally-anchored stereotypes, and they are often applied without the perceivers' intention or conscious awareness (Macrae & Bodenhausen 2000). This last point relates to the third aspect of person schemata. Person schemata are unities of characteristic features integrated at the level of persons. All these singular features are integrated into person models that enable us to develop detailed and extensive expectations of behaviour.

(iii) *Integration of characteristic features at the level of perceiving the whole person*: Although I have presented evidence that some single features are very salient for transferring social information, there is also much evidence that these features are normally combined with a variety of others to form an integrated impres-

13 We leave the question open as to what extent person schemata are constituted by innate or by learned dispositions. The examples mentioned above indicate that they involve properties of both kinds.

14 Interestingly, Germans could perceive the friendliness of students from the US and UAE partially (as well as the other way around), while students from UAE and USA could not read the level of friendliness from the other culture at all (Bente et al. 2010).

sion of a person that I call a person schema. We have seen evidence for the key role of gaze detection in registering another person's direction of attention (see ii). But there is further evidence that gaze alone is not the critical source of information; we actually seem to rely on an integrated evaluation on the basis of perceiving gaze, head, and body position (Frischen et al. 2007). The same holds for evaluation of the basic features sex, race, and age. Although isolated facial features are often sufficient to determine a person's sex, research has indicated that sex categorization is based on the integration of several features (Baudoin & Humphreys 2006; Bruce et al. 1993; Brown & Perrett 1993; Roberts & Bruce 1988; Schyns et al. 2002). Concerning face, the best available theory of face recognition seems to be Haxby's account (Haxby et al. 2000), according to which there are two distinguishable processes, one leading to face identification by focussing more on invariant core features, and the other leading to registering facial expression by relying on varying features. Furthermore, there is evidence that there are two different neural circuits for face perception and body perception (see the review by Macrae & Quadflieg 2010), both playing a core role in registering face or body identity, and playing an extended role in registering face or body expression in a given situation. And the integration processes are not limited to this level (Martin & Macrae 2007). Since we know that information about facial and bodily features is integrated, e.g., in the evaluation of emotional expression, we can therefore characterize a sequence of integration processes as leading finally to a person impression in a situation, which may be stored as a person schema in memory.

5.2 A model of forming a person schema

How can we best describe this process of forming a person schema? In general terms, the same complex process takes place in the case of perceiving a person and forming a person impression in a given situation as takes place when we perceive an object. I describe the process according to the model of object perception developed by Ernst & Bülthoff (2004), and I have already shown in detail that it can do justice to our recognition of emotions (Newen et al. forthcoming). The overall process comprises bottom-up processes starting with basic visual features that are modulated either by feature combination (if two features provide complementary information), or by feature integration. The latter can be modelled as a Bayesian weighting process that leads to the most probable intermediate estimate given the input. Further integration processes then lead from the most probable estimate to a stable percept of an object in the case of object perception, and to a stable person impression in the case of person perception. This model explicitly accounts not only for bottom-up but also for top-down processes, in the form of so-called cognitive penetration. I have sketched a plausible but in no way complete model of the formation of a person impression (see figure below). According to the evidence I have presented so far, it is plausible to suggest that at the level of intermediate estimates in the process of forming an impression of a person, we find (a) an estimation of a core person identity, (b) an estimate of situational emotions, intentions, and actions, as well as (c) an estimation of social status, person abilities, and individual personality traits. An important step in the model is the association of visual features with socially-anchored stereotypes (see above) which allows us to develop rich intermediate estimates, e.g., of the other's emotional situation, social status, etc.

> Numerous lines of research (Albright, Kenny, & Malloy, 1988; Ambady & Rosenthal, 1992; Behling & Williams, 1991; Borkenau & Liebler, 1992; Kenny, Horner, Kashy, & Chu, 1992; Norman & Goldberg, 1966; Secord, Dukes, & Bevan, 1954) have provided compelling evidence that trait evaluations are readily drawn from a person's physiognomy (i.e., facial features), outer appearance (i.e., clothing), or demeanor (i.e., posture, walking, style). (Macrae & Quadflieg 2010, p. 433)

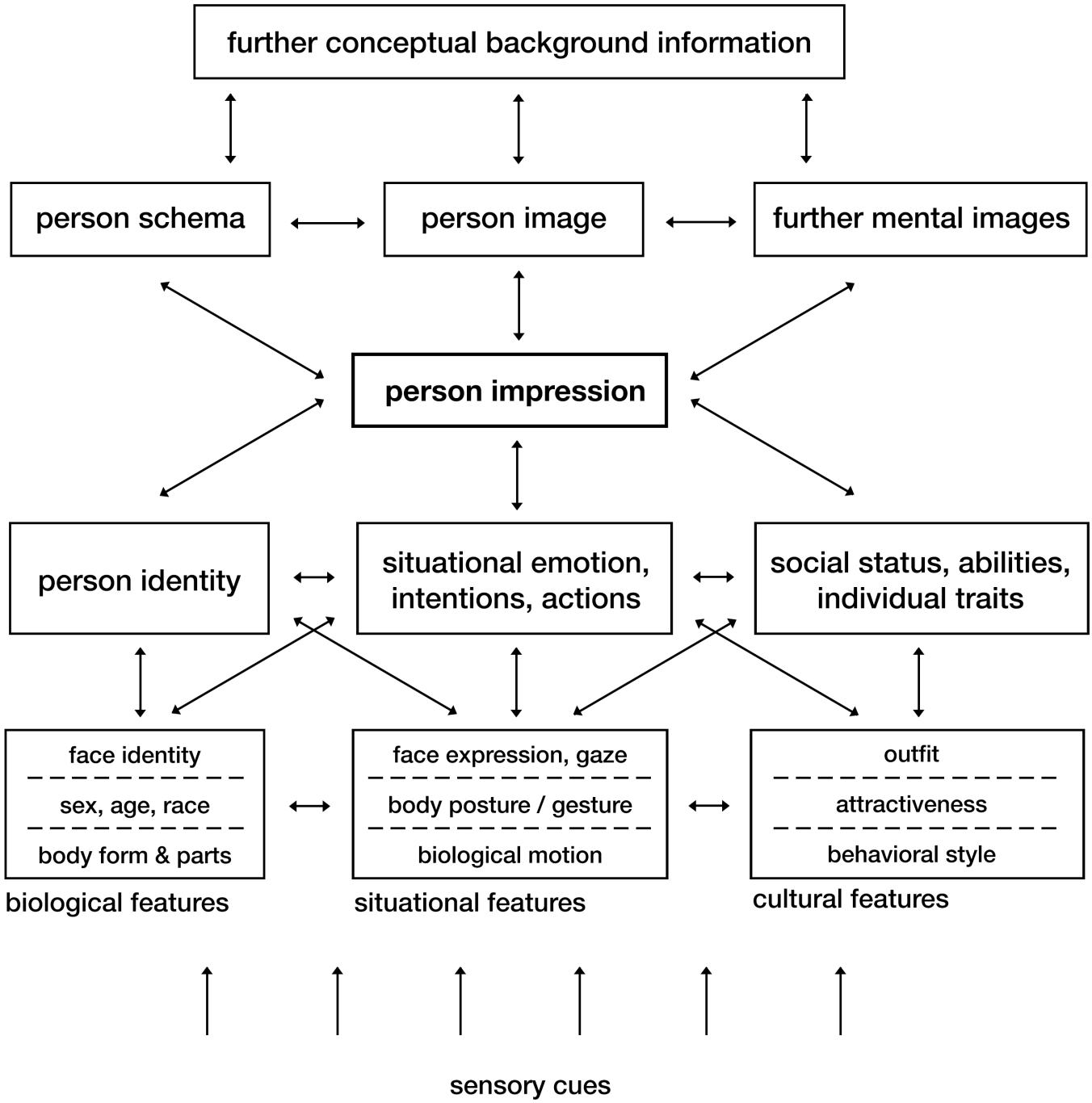

Figure 1: A model of the dynamics of bottom-up and top-down processes leading to a stable person impression by relying on person images and/or person schemata

Finally, I highlight that the top-down processes are able to interfere in this process of combination and integration very early in the visual information processes: for example, it has been shown that the activation of a race concept on the basis of the form of a face (African versus European face format) changes the perception of colour in the face, while colour is known to be represented in V4 as part of early visual brain processes. The same hue of colour is seen as more dark in the African face than in the European face (Levin & Banaji 2006). Thus we have to admit that the process of feature-combination and integration is highly dynamic, involving simultaneous activation of features rooted in bottom-up and top-down processes, finally reaching the most probable and usually stable person impression. The dynamic is de-

scribed in detail for the case of object perception in Vetter & Newen (2014); it is postulated for person categorization in Macrae & Martin (2007), and analysed according to the levels of processing that lead to person construal in Freeman & Ambady (2011). Figure 1 is a sketch of the formation of a person impression according to my account.

A person schema emerges as the result of direct perception of a person, where this may be either basic or relatively smart perception; yet it usually remains implicit, and is not amenable to linguistic description. A typical example of person schema based on basic perception is the everyday experience of seeing a person only briefly in a single situation, whereupon it is difficult for us to describe the person—particularly her face. While we can often easily recognize the person, it may take hours with a professional to end up with an adequate "identikit" picture such as those produced at police stations. A person schema based on smart perception might be, for instance, a person schema that includes a lot of top-down activation—for example, while on campus, perhaps I see a person of typical student age dressed like a law student, and thus activate the "rich person" schema that is the basis for my everyday smooth interaction with law students, and which differs (despite overlaps) from my person schema for students in natural sciences. If we not only develop implicit practical knowledge regarding our use of the person impression (independent from its richness), but also develop explicit *knowledge* pertaining to the relevant person information, or at least develop easy explicit access to it, then we go beyond a person schema. We can characterize this new unified information as a person image.

5.3 Person images

In detail, then, what is a person image? A person image is a unity of relatively easily and explicitly available information about a person, including her mind-set. On the basis of typically implicit person schemata, young children learn to develop explicit *person images*. These are models of individual subjects or groups. In the case of individual subjects, they may include names, descriptions, stories, whole biographies, and visual images highlighting both mental and physical dispositions as well as episodes. Person images are essentially developed not only by observation but also by telling, exchanging, and creating stories (or "narratives").[15] Person images presuppose the capacity to explicitly distinguish the representation of my own mental and physical phenomena from the representation of someone else's mental and physical phenomena. This ability develops gradually, reaching a major and important stage when children acquire the so-called explicit theory-of-mind ability (operationalized by the false-belief task, see Wimmer & Perner 1983).[16] Then they are able to construct explicit person images by characterizing a person such that they attribute a biography to an individual. There is strong folk-psychological evidence that we have explicit person models of the people we deal with extensively, e.g., family members, and people about whom we tend to have a lot of explicit knowledge. The same is true for relevant groups of persons we deal with often. Even in professional contexts this leads to judgments that can be inadequate: the apparent association between wearing revealing clothes and immodesty and promiscuity has been shown to cause not only laypeople but also police officers and judges to hold victims of rape to be responsible for their having been assaulted (Lennon et al. 1999). An essential part of becoming an adult is learning to interact socially with other humans, by developing sophisticated and explicit person images of the groups of professions we have to come to any sort of arrangement with. We often have explicit beliefs about medical doctors, managers, secretaries, craftspeople, etc., and we try to deploy these beliefs to deal with these people in a smooth and efficient way. When we

15 This is the aspect of the narrative approach to understanding other minds, mentioned above (e.g., Hutto 2008). But narratives are only one method of establishing a person model. Representatives of a pure narrative approach underestimate the importance of other sources, such as perceptions, feelings, interactions, etc., which often do not involve narratives.

16 There is a long and not fully understood process of development from implicit false belief sensitivity to explicit false belief understanding (de Bruin & Newen 2012a; 2012b). Person images actually presuppose an explicit representation of false beliefs.

have stored a person image in memory, and are placed in a new situation in which we see and recognize the person, there is evidence that we immediately activate the biographical knowledge we have available. For example, when test persons were asked to judge the traits of target individuals from photographs, the test persons' responses continue to be influenced by what they have explicitly learned about the people in question (Uleman et al. 2005). A recent neuroimaging study (Hassabis et al. 2013) indicated that when test persons were asked to predict the behaviour of persons, they essentially relied on prior knowledge of personality traits, which in this particular study were implemented in two ways, namely as agreeableness (the tendency toward altruism, cooperation, and the valuing of harmony in interpersonal relationships as opposed to antisocial and exploitative behaviours) and as extroversion (in contrast to introversion). The test person became acquainted with four types of personalities that had been constructed from combinations of high and low versions of agreeableness, on the one hand, and high and low versions of extroversion, on the other. In the test situation they had to predict the behaviour of four specific persons who were exemplars of the four personality types. The authors report that the predictions of behaviour were mainly based on personality traits and that the latter also had rather clear neural correlates: by using functional magnetic resonance imaging (fMRI) the authors showed that there is a neural correlate for recognizing (and imagining) high agreeableness (in contrast to low), namely in the left LTC (lateral temporal cortex) and dorsal mPFC (medial prefrontal cortex), as well as for recognizing (and imagining) high extroversion (in contrast to low), namely in the pCC (posterior cingulate cortex); in addition the recognition (and imagination) of one of the four personality types was correlated with four distinctive patterns in the anterior medial prefrontal cortex (mPFC). In line with my proposal, the authors of the fMRI study write: "Different patterns of activation in the anterior mPFC could reliably distinguish between the different people whose behavior was being imagined. It is hypothesized that this region is responsible for assembling and updating personality models" (Hassabis et al. 2013). Since the study was based on explicit evaluation of personality features or types, I take this to support the existence of person images. Yet even if the reader accepts the idea of person models, she may be sceptical about whether we need to distinguish person *schemata* and person *images*.

5.4 Why should we distinguish person schemata and person images?

A very convincing case that forces us to make a distinction between person schemata and person images comes from taking a closer look at a typical patient suffering from Capgras syndrome, a misidentification syndrome. Sufferers have the delusional belief that one of their closest relatives, e.g., their wife, has been replaced by an impostor. Such a patient typically says things like "this person looks exactly like my wife, she even speaks and behaves like my wife and she expresses her typical desires but she is not my wife" (Davies et al. 2001); thus, one aspect of this mental disorder is the observation that all the features explicitly believed to be possessed by the wife are correctly attributed. We can account for this by asserting that the patient has an intact person image of his wife. Nevertheless, the usual person identification has gone wrong. According to a standard analysis, what is lacking in the case of the Capgras patient is a feeling of familiarity that normally comes with perceiving a well-known person. How can we account for this in the new framework? When perceiving his wife, the subject intuitively develops and activates a person schema. One aspect of the person schema is the person's identity.[17] As the Capgras case nicely illustrates, the registration of a person's identity is a result of an integration process that relies not only on visual features but also on an implicit emotional evaluation, and that these together trigger an explicit judgment. While the

17 The involvement of identity already at the level of implicit schemata is supported by Haxby's model of face perception according to which we have to distinguish a core cognitive system involving the recognition of face identity and an extended cognitive system which is enabling the recognition of facial expression (Haxby et al. 2000).

visual recognition fits, here the emotional evaluation is inadequate and the feeling of familiarity is lacking; and in the case of this disorder, the Bayesian integration process for these features leads to an implausible result, since the emotional mistake overrides the visual adequateness. Thus, the Capgras patient has an adequate person image of his wife but an incorrect person schema, and the tension between the two is solved by developing the (implausible) hypothesis that she is an imposter. This analysis is in line with two-factor theories of the Capgras disorder, according to which two distinct factors cause the phenomenon[18]: first, the lack of familiarity, and, second, a local breakdown of rationality that enables the irrational belief-formation on the basis of a severely disturbed person schema (Davies et al. 2001).[19] Several other cases seem to be accounted for if we accept the evidence for a two-factor theory of person modelling—namely a first level of intuitive and implicit person impression and a second level of explicit person evaluation, which are described respectively as intuitive person schemata and explicit person images.

A contrast case to Capgras syndrome is the Fregoli syndrome, wherein a patient has the delusional belief that one and the same person, usually a persecutor, is following her, who is able to radically change his outer appearance. The sufferer then connects people with rather different outer appearances and treats them as the same persecutor. One explanation, still in need of testing, is that this time the feeling of familiarity is developed too often, probably by top-down initiation due to the delusional belief that the subject is being persecuted. The delusional belief, together with an inadequate feeling of familiarity, may explain the syndrome.[20] But again we need to distinguish the two factors: a level of implicit feeling or impression, and a level of explicit judgment. This time the delusion produces a breakdown of rational judgment formation, i.e., the person model of the other is strongly influenced by the delusion: the person schema formation may be largely intact but has a local defect due to being dominated by the delusional belief. In general, monothematic delusions (delusions about a single belief content) seem to rely on two factors (Coltheart et al. 2007): "[o]ne factor has to explain the strange experiences patients claim to have, while the other factor has to explain the misattribution of actions and thoughts" (Vosgerau & Newen 2007, p. 40).

Are there nonpathological everyday cases that support the distinction between person schema and person image? One illustration can be drawn from Mark Twain's "Huckleberry Finn." At first Huck helps the slave Jim to escape from slavery; but then he rethinks his support in the light of the law, and forms the judgment that he should turn him in to the slave-hunters. But when he has the opportunity to do so, Huck actually ends up protecting Jim. Why does he do this? Huck has a person schema of Jim that is constituted by a person impression according to personal interactions that are dominated by empathy; thus he has a positive impression of Jim and there exists between them a growing friendship. On the other hand, he has a person image of Jim that is dominated by the fact that he is a slave, such that he has to accept his role in society, to do the hard work, to live without freedom, and thus that it is forbidden to aid his escape. Cases of tension between an intuitive person impression (being helpful, being peaceful) and a person image dominated by the knowledge that the same person is a pathological murderer are often reported by judges and policemen. A less dramatic tension seems to be part of our everyday experience of "false" friends (we may still think of someone as

18 In the literature there are discussed one-factor accounts to explain mental disorders, e.g., in the case of schizophrenia (Gallagher 2004): a top-down approach argues that disturbances of higher-order cognition is the only source for thought insertion (Stephens & Graham 2000) while a bottom-up approach argues that thought insertation is a product of disturbances of neural or basic cognitive processes (like perception). Most of the recent accounts are hybrid account which we call two-factor theories.

19 The fact that person identity as a component of person schema formation is not only based on visual but also on an emotional evaluation is supported by the case of prosopagnosia, i.e., the inability to recognize the face of the person one is seeing, even though one is able to see and perceive the rest of the person adequately. Despite the fact that a person suffering from prosopagnosia is not able to see the familiarity of the face, we can measure increased skin conductance for familiar but not unfamiliar faces, thereby demonstrating intact (albeit covert) emotional recognition of known others (de Haan et al. 1992; Tranel & Damasio 1985).

20 For a discussion of delusional phenomena, see Coltheart et al. (2007) and Hirstein (2005).

a friend while implicitly already noticing signs of unfair treatment), though of course the tension can also exist the other way around. As illustrated above, the visual features of a person are often loaded with social information, and often involve the activation of negative prejudices which, after a more careful investigation of the person, can be opposed by a positive person image. The general functional role of person models is to simplify the structuring and evaluation of social situations, to enable a quick evaluation of the person in a given situation, and to initiate adequate behaviour. An additional special functional role of person models consists in stabilizing my self-estimation, since there is a strong tendency to have positive stereotypes of one's own in-group members and negative stereotypes of the out-groups' (see Volz 2008, p. 19). These examples illustrate not only that we need to distinguish the person schema and person image, but also that we have a tendency towards harmonizing both. Thus, if one of them is disturbed we tend to adjust the other, which may result not only in wrong judgments about persons, but in extreme cases may become an aspect of a mental disorder, as described above. Finally, to distinguish them is compatible with the claim that a person image may often gradually evolve on the basis of a person schema such that partially the same information about a person changes the status of accessibility from implicit to explicit. But we also have to distinguish both kinds of person models because often an implicit representation of a person as unfriendly exists simultaneously with an explicit evaluation of the same person as friendly.

5.5 Person model theory (PMT) and its relation to other main theories

The central claim of PMT is that we organize information about others by forming person models. We account for a multiplicity of epistemic access strategies, while direct perception and interaction are the main source for person schema formation. Person image formation is based on all the epistemic strategies we have examined, including theory-based inferences and (high-level) simulation strategies. Why, then, is PMT not a version of TT? Person models are more general and allow for a unification of rather parsimonious information about a person, which does not warrant being called a theory since it does not form even a minimal package of systematically-interconnected beliefs. As we learn more and more about the same person, our person model may develop into a theory. Thus, this is not to deny that we often have rich person models that are theories; and thus I can account for the empirical evidence that supporters of TT tend to rely on. A further question concerns how PMT is related to ST. Simulation is one epistemic strategy in which person models are used to understand others: if I have evidence that another person is similar to me in relevant respects, then I may use my self-model, either the self-schema or the self-image, to produce an explanation or a prediction of the other's behaviour. But I also often have clear knowledge that the other is different from me in relevant respects, especially when there are great differences in the three main categories—sex, age, and race—or in cultural background. In such cases simulation is not used. Although simulation is a worthy epistemic strategy, it is only of limited and constrained use in everyday understanding. How is PMT related to interaction theory and direct perception theories? It explicitly accepts the important role of both as epistemic strategies, but insists that in addition to understanding others in situations of direct interaction there is also often an understanding of others just by observation. The use of these two strategies seems to depend heavily on the personality traits of the person who aims to understand another: while extroverts mainly rely on interaction, introverts (who avoid social contact) mainly rely on observation. Furthermore, these theories do not offer an answer to the main question addressed in this article, namely how we organize the information about other people that we already have. The narrative account offers one answer here, and again we can account for the role of narratives that in the case of rather rich person models may be sources for creating or enriching the models further, or they may also concern the way a person model is memorized. But the narrative account

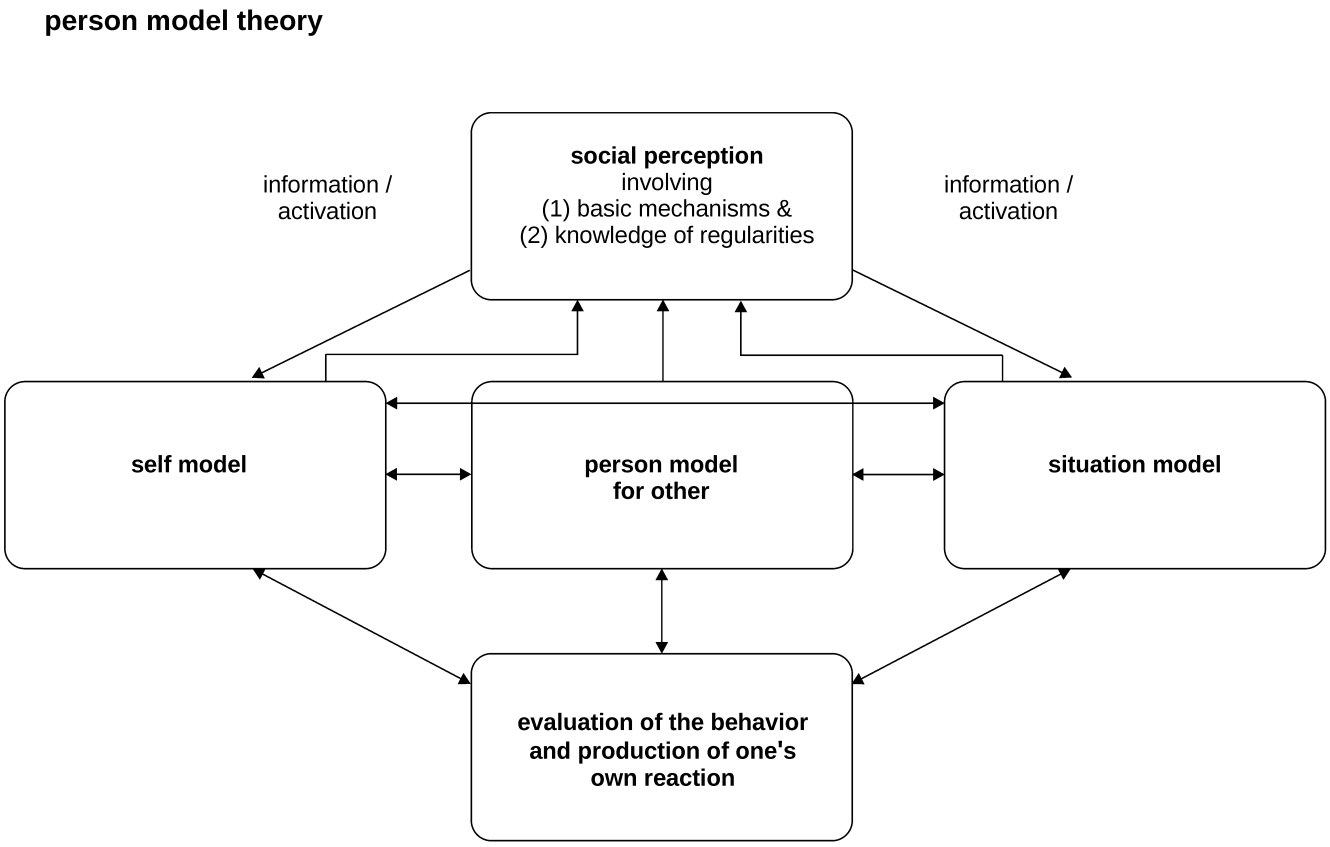

Figure 2: Interaction of person models with situation models in understanding others

alone ignores the strong relevance of our intuitive understanding of others as it is anchored in person schemata. This short overview, then, indicates that all of the evidence that representatives of other theories put forward can be integrated into this view, while there is further evidence for my theory, e.g., rich evidence that there is an integration of information into person models by person perception. Notably, PMT allows us to account for certain mental disorders, and I have cited evidence from a very recent fMRI study that is further supportive of the organization of information according to person models.

5.6 Widening PMT: Person models, situation models and culture

Does PMT give us the complete story about understanding others? What about my understanding of a person whom I only see from behind, when queuing at a self-service restaurant? Here it seems sufficient to predict her behaviour just by expecting her to act according to the social conventions of a self-service restaurant. Understanding the situation alone seems to be sufficient for an understanding of and interaction with the other.[21] This is an important observation that suggests a widening of my theory: we do not only create person models, but also situation models, and our understanding of others uses both types of model as input and selects the most helpful model for evaluating the other person. If I have no person model of this individual, if seeing someone from the back gives me only very parsimonious information, and if I am only interested in getting my lunch, then the situation model may be dominant in dealing with persons in this context. As soon as minimal enrichment of person information is available we naturally tend to rely on person models. The fact that situation models are used at all is supported by successful artificial intelligence (AI) studies working with scripts and

21 These types of cases are considered in Gallagher & Hutto (2008), in the section "Pragmatic Intersubjectivity". Their view is close to a multiplicity view. A minor criticism is that we have to account for such cases independently from being in interaction with someone. They may also involve only observing the other.

frames that can account for human behaviour (Schank & Abelson 1977). Furthermore, in Asian cultures the understanding of other people seems to rely much more on social conventions, since people are strongly expected to behave according to these conventions. In general, situation models are more important for understanding others in "collectivistic" cultures than in individualistic cultures where explanations and predictions of behaviour are usually more reliant on individual belief–desire explanations. Such observations as these require us to give an account of situation models. This can be easily done by widening the theory of understanding others such that it includes situation models, as well as the interdependence of personal models and situation models. It can also include a dynamic, involving bottom-up and top-down processes that lead to an activation or construction of the most plausible person model for interacting with, explaining, or predicting the behaviour of the other person.[22] Here is a rough outline of the process leading to understanding others in the rich sense of interacting, such as in observing, explaining, or predicting (see Figure 2).

In general, we should note the important role of culture in shaping our way of modelling persons (Vogeley & Roepstorff 2009). As we have seen, culture modulates the relevance of person models in relation to situation models. But it also influences our formation of person models, for example by shaping our person perception. To illustrate: Japanese individuals are encouraged to be sociable and cooperative (Moskowitz et al. 1994), to be affiliative rather than competitive (Yamaguchi et al. 1995), and to show obligation to others (Oyserman et al. 1998). Concerning dominance and subordination, Japanese people learn to be rewarded for subordinate behaviour, while Americans learn to be rewarded for dominant behaviour. This also shapes the perception of dominance and subordination in others. Typical neurological activations of the mesolimbic reward system can be shown to be shaped by the respective culture: Americans show a higher activation of this system when doing and seeing dominant behaviour (in contrast to subordinate behaviour) while with Japanese people we can observe the opposite: they show a higher activation of exactly the same system when doing and seeing subordinate behaviour (Freeman et al. 2009). Thus, the perception of dominant and subordinate behaviour is connected with opposite evaluations (Americans highly esteem dominance while Japanese people highly esteem subordinate behaviour) and a different set of personality traits. Cultural influences on the psychological and neural level are also reported for self-models: on the psychological level, the difference between an Asian interdependent self and a Western independent self was reported by Markus & Kitayama (1991), while a respective difference in neural correlates was also recently discovered (Sui & Han 2007).

6 Conclusion

Our understanding of other minds is based epistemically on a multiplicity of strategies, the core strategies being direct perception, interaction, simulation, and theory-based inferences (including learning from narratives). The most important aspect of understanding others is the activation of prior knowledge of individuals or groups of persons. This is organized into person models. The main claim of PMT is that we rely on *person models* to understand others. These person models form the basis for perceiving and evaluating persons, their social behaviour, and their mind-set. We develop person models for ourselves, for other individuals, and for groups of persons (group models). Furthermore, all types of person models can be realized on two levels: (implicit) person schemata and (explicit) person images. A *person schema* is a bundle of information including information about sensory-motor abilities, voice, face, basic mental dispositions, etc., and such schemata are intuitively used, implicitly developed, and not usually

22 There is already one dynamic model of person construal available in the literature that also supports my dynamic theory of understanding others with person models, i.e., the model of Freeman & Ambady (2011). Despite its merits in describing social perception in more detail as regards the interrelation of bottom-up and top-down processes, the authors neither account for the claim that our rich prior information is mainly organized on the level of persons (not faces or subpersonal features), nor do they account for the interaction between person models and situation models.

easily accessible for linguistic report. A *person image* is a unity of explicitly-registered mental and physical dispositions as well as situational features (like perceptions, emotions, attitudes, etc.) that is usually easily accessible for linguistic report (albeit sometimes with the help of gesture, drawings, etc.). The PMT has several advantages over existing accounts of social understanding (e.g., TT, ST, and interaction theory), since it can account for all of the following criteria:

1. It explains specific and more general social understanding of particular individuals in terms of individual person models and group person models. (Not accounted for in ST.)
2. It accounts for the difference, for which evidence is presented, between implicit, intuitive forms of social understanding and explicit deliberative ones by appealing to the role of person schemata and person images respectively. (Not accounted for in interaction theory.)
3. It does justice to folk-psychological evidence that we understand very familiar persons much better than unfamiliar ones: We have rich person images of individuals with whom we are very familiar. (Deficit of all former theories.)
4. It marks adequately in what ways our understanding of others and our self-understanding are interdependent, e.g., in special cases of simulation, understanding the other relies on self-models. (Generally not accounted for in TT.)
5. It offers an adequate framework that is in line with the best explanations of some mental diseases in understanding others, such as the Capgras and Fregoli syndromes. (Deficit of ST.)
6. It can account for cultural differences in social understanding: Future research will show how person models vary with culture, and we have already illustrated that it varies in the case of self-models between Asian and Western cultures. (Not accounted for in any former theory.)

Thus, PMT is at least a serious alternative account, and certainly a candidate for future investigation.

Acknowledgements

I wish to express special thanks to Luca Barlassina, Kai Vogeley, Anna Welpinghus, and Tobias Starzak for their helpful comments. This paper is part of the project "Social Information Processing and Culture" funded by the VolkswagenStiftung.

References

Adams, R. B., Jr. & Kleck, R. E. (2003). Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science*, *14* (6), 644-647.

Albright, L., Kenny, D. A. & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, *55* (3), 387-395. 10.1037/0022-3514.55.3.387

Ambady, N. & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111* (2), 256-274. 10.1037/0033-2909.111.2.256

Bar, M., Neta, M. & Linz, H. (2006). Very first impressions. *Emotion*, *6* (2), 269-278. 10.1037/1528-3542.6.2.269

Baron-Cohen, S. (1995). *Mindblindness. An essay on autsim and theory of mind.* Cambridge, MA: MIT Press.

Baudoin, J.-Y. & Humphreys, G. W. (2006). Configural information in gender categorisation. *Perception*, *35* (4), 431-450. 10.1068/p3403

Behling, D. U. & Williams, E. A. (1991). Influence of dress on perception of intelligence and expectations of scholastic achievement. *Clothing and Textiles Research Journal*, *9* (4), 1-7. 10.1177/0887302X9100900401

Bente, G., Leuschner, H., Al Issa, A. & Blascovich, J. J. (2010). The others: Universals and cultural specificities in the perception of status and dominance from nonverbal behavior. *Consciousness and Cognition*, *19* (3), 762-777. 10.1016/j.concog.2010.06.006.

Bentin, S. & Deouell, L. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanism. *Cognitive Neuropsychology*, *17* (1-3), 35-54.

Bentin, S., Allison, T., Puce, A., Perez, E. & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8* (6), 551-565. 10.1162/jocn.1996.8.6.551

Bertin, E. & Striano, T. (2006). The still-face response in newborn, 1.5-, and 3-month-old infants. *Infant Behavior and Development*, *29* (2), 294-297. 10.1016/j.infbeh.2005.12.003.

Borkenau, P. & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62* (4), 645-657. 10.1037/0022-3514.62.4.645

Brewer, M. B. (1988). A dual-process model of impression formation. In R. S., Jr. Wyer & T. K. Srull (Eds.) *Advances in Social Cognition* (pp. 1-36). Mahwah, NJ: Erlbaum.

Brown, E. & Perrett, D. I. (1993). What gives a face its gender? *Perception*, *22* (7), 829-840. 10.1068/p220829

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R. & Linney, A. (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception*, *22* (2), 131-152. 10.1068/p220131

Buckley, S. J., Bird, G. & Sacks, B. (2002). Social development for individuals with down syndrome: An overview.

Calder, A. J., Keane, J., Manes, F., Antoun, N. & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience*, *3*, 1077-1078. 10.1038/80586

Carroll, J. M. & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology*, *70* (2), 205-218. 10.1037/0022-3514.70.2.205.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32* (2), 121-182. 10.1017/S0140525X09000545

Cleveland, A. & Striano, T. (2007). The effects of joint attention on object processing in 4- and 9-month-old infants. *Infant Behavior and Development*, *30* (3), 499-504. 10.1016/j.infbeh.2006.10.009

Coltheart, M., Langdon, R. & McKay, R. (2007). Schizophrenia and monothematic delusions. *Schizophrenia Bulletin*, *33* (3), 642-647. 10.1093/schbul/sbm017.

Davies, M., Coltheart, M., Langdon, R. & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy,Psychiatry and Psychology*, *8* (2/3), 133-158. doi: 10.1353/ppp.2001.0007.

de Bruin, L., van Elk, M. & Newen, A. (2012). Reconceptualizing second-person interaction. *Frontiers in Neuroscience*, *151*, 1-10. 10.3389/fnhum.2012.00151

de Bruin, L. & Newen, A. (2012a). An association account of false belief understanding. *Cognition*, *123* (2), 240-259. 10.1016/j.cognition.2011.12.016.

——— (2012b). The developmental paradox of false belief understanding: A dual-system solution. *Synthese*, *191* (3), 297-320. 10.1007/s11229-012-0127-6

de Haan, E. H. F, Bauer, R. M. & Greve, K. W. (1992). Behavioural and physiological evidence for covert face recognition in a prosopagnosic patient. *Cortex*, *28* (1), 77-95. 10.1016/S0010-9452(13)80167-0

Devine, P. G. & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited.

Personality and Social Psychology Bulletin, *21* (11), 1139-1150. 10.1177/01461672952111002

Dion, K., Berscheid, E. & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24* (3), 285-290. 10.1037/h0033731

Eagly, A. H., Ashmore, R. D., Makhijani, M. G. & Longo, L. C. (1991). What is beautiful is good, but …: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110* (1), 109-128. 10.1037/0033-2909.110.1.109

Eimer, M. & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *NeuroReport*, *13* (4), 427-431. 10.1097/00001756-200203250-00013

Eimer, M. & McCarthy, R. A. (1999). Prosopagnosia and structural encoding of faces: Evidence from event-related potentials. *NeuroReport*, *10* (2), 255-259. 10.1097/00001756-199902050-00010

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.) *Nebraska Symposium on Motivation, 1971, Vol. 19* (pp. 207-283). Lincoln, NE: University of Nebraska Press.

——— (1999). Basic emotions. In T. Dalgleish & M. J. Power (Eds.) *The Handbook of Cognition and Emotion* (pp. 45-60). New York, NY: Wiley.

Ekman, P., Friesen, W. V. & Ellsworth, P. (1972). *Emotion in the Human Face.* New York, NY: Pergamo.

Ernst, M. O. & Bülthoff, H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8* (4), 162-168. 10.1016/j.tics.2004.02.002

Fiebich, A. & Coltheart, M. (under review). *Various ways to understand other minds.*

Fiske, S. T. & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1-74. 10.1016/S0065-2601(08)60317-2

Freeman, J. B. & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118* (2), 247-279. 10.1037/a0022327

Freeman, J. B., Rule, N. O. & Ambady, N. (2009). The cultural neuroscience of person perception. *Progress in Brain Research*, *178*, 191-201. 10.1016/S0079-6123(09)17813-5

Frey, S. (1999). *Die nonverbale Kommunikation.* Bern, SUI: Huber.

Frischen, A., Bayliss, A. P. & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133* (4), 694-724. 10.1037/0033-2909.133.4.694

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

——— (2004). Neurocognitive models of schizophrenia: A neurophenomenological critique. *Psychopathology*, *37* (1), 8-19. 10.1159/000077014

——— (2005). *How the body shapes the mind.* Oxford, UK: Oxford University Press.

——— (2007). Simulation trouble. *Social Neuroscience*, *2* (3), 353-365. 10.1080/17470910601183549.

——— (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17* (2), 535-543. 10.1016/j.concog.2008.03.003

Gallagher, S. & Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. Racine, C. Sinha & E. Itkonen (Eds.) *The shared mind: Perspectives on intersubjectivity* (pp. 17-38). Amsterdam, NL: John Benjamins.

Gliga, T. & Dehaene-Lambertz, G. (2005). Structural encoding of body and face in human infants and adults. *Journal of Cognitive Neuroscience*, *17* (8), 1328-1340. 10.1162/0898929055002481

Gobet, F. (1997). Roles of pattern recognition and search in expert problem solving. *Thinking and Reasoning*, *3* (4), 291-313. 10.1080/135467897394301

Gobet, F. & Simon, H. A. (1996). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, *3* (2), 159-163. 10.3758/BF03212414

Goldman, A. I. (2006). *Simulating minds. The philosophy, psychology, and neuroscience of mindreading.* Oxford, UK: Oxford University Press.

Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16* (1), 1-14. 10.1017/S0140525X00028636

Gopnik, A. & Meltzoff, A. (1997). *Words, thoughts, and theories.* Cambridge, MA: MIT Press.

Griffiths, P. E. (1997). *What Emotions Really Are. The Problem of Psychological Categories.* Chicago, IL: Chicago University Press.

Hassabis, D., Spreng, R. N., Rusu, A. A., Robbins, C. A., Mar, R. A. & Schacter, D. L. (2013). Imagine all the people: how the brain creates and uses personality models to predict behavior. *Cerebral Cortex, March 5.* 10.1093/cercor/bht042

Haxby, J. V., Hoffman, E. A. & Gobbini, M. A. (2000). The distributed human neural system for face percep-

tion. *Trends in Cognitive Sciences*, *4* (6), 223-233. 10.1016/S1364-6613(00)01482-0

Heberlein, A. S., Adolphs, R., Tranel, D. & Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from pointlight walkers. *Journal of Cognitive Neuroscience*, *16* (7), 1143-1158. 10.1162/0898929041920423

Herrmann, M. J., Ehlis, A. C., Muehlberger, A. & Fallgatter, A. J. (2005). Source localization of early stages of face processing. *Brain Topography*, *18* (2), 77-85. 10.1007/s10548-005-0277-7

Hirstein, W. (2005). *Brain fiction. Self deception and the riddle of confabulation.* Cambridge, MA: MIT Press.

Holmes, A., Vuilleumier, P. & Eimer, M. (2003). The processing of emotional facial expression is gated by spatial attention: Evidence from event-related brain potentials. *Cognitive Brain Research*, *16* (2), 174-184. 10.1016/S0926-6410(02)00268-9

Hutto, D. (2008). *Folk-psychological narratives.* Cambridge, MA: MIT Press.

Itier, R. J. & Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex*, *14* (2), 132-142. 10.1093/cercor/bhg111

Kenny, D. A., Horner, C., Kashy, D. A. & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology*, *62* (1), 88-97. 10.1037/0022-3514.62.1.88

Kuzmanovic, B., Schilbach, L., Lehnhardt, F. G., Bente, G. & Vogeley, K. (2011). A matter of words: Impression formation in complex situations relies on verbal more than on nonverbal information in high-functioning autism. *Research in Autism Spectrum Disorders*, *5*, 604-613.

Lennon, S. J., Johnson, K. K. P. & Schulz, T. L. (1999). Forging linkages between dress and law in the U.S., part I: Rape and sexual harassment. *Clothing and Textiles Research Journal*, *17* (3), 144-156. 10.1177/0887302X9901700305

Levin, D. T. & Banaji, R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, *135* (4), 501-512. 10.1037/0096-3445.135.4.501

Lin, M. H., Kwan, V. S. Y., Cheung, A. & Fiske, S. T. (2005). Stereotype content model explains prejudice for an envied outgroup: Scale of Anti-Asian American stereotypes. *Personality and Social Psychology Bulletin*, *31* (1), 34-47. 10.1177/0146167204271320

Liu, J., Harris, A. & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5* (9), 910-916. 10.1038/nn909

Macpherson, F. (2012). Cognitive penetration of colour experience. Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, *84* (1), 24-62. 10.1111/j.1933-1592.2010.00481.x.

Macrae, C. N. & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93-120. 10.1146/annurev.psych.51.1.93

Macrae, C. N. & Martin, D. (2007). A boy primed Sue: Feature-based processing and person construal. *European Journal of Social Psychology*, *37* (5), 793-805. 10.1002/ejsp.406

Macrae, C. N. & Quadflieg, S. (2010). Perceiving people. In S. Fiske, D. T. Gilbert & G. Lindzey (Eds.) *Handbook of social psychology* (pp. 428-463). New York, NY: McGraw-Hill.

Markus, H. R. & Kitayama, S. (1991). Culture and the self. Implications for cognition, emotion, and motivation. *Psychological Review*, *98* (2), 224-253. 10.1037/0033-295X.98.2.224

Martin, D. & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology*, *37* (5), 37-5. 10.1002/ejsp.445

Meeren, H. K. M., van Heijnsbergen, C. C. R. J. & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America*, *102* (45), 16518-16523. 10.1073/pnas.0507650102

Meltzoff, A. N. & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198* (4312), 75-78. 10.1126/science.198.4312.75

——— (1994). Imitation, memory and the representation of persons. *Infant Behaviour and Development*, *17* (1), 83-99. 10.1016/0163-6383(94)90024-8

Moskowitz, D. S., Suh, E. J. & Desaulniers, J. (1994). Situational influences on gender differences in agency and communion. *Journal of Personality and Social Psychology*, *66* (4), 753-761. 10.1037/0022-3514.66.4.753

Nagy, E. (2008). Innate intersubjectivity: Newborn's sensitivity to communication disturbance. *Developmental Psychology*, *44* (6), 1779-1784. 10.1037/a0012665

Nelson, C. A. (2001). The development and neural bases of face recognition. *Infant and Child Development*, *10*

(1-2), 3-18. 10.1002/icd.239

Newen, A., Welpinghus, A. & Juckel, G. (forthcoming). *Emotion recognition as pattern recognition: the relevance of perception.*

Newen, A. & Schlicht, T. (2009). Understanding other minds. A criticism of Goldman's simulation theory and an outline of the person model theory. *Grazer philosophische Studien*, *79* (1), 209-242.

Newen, A. & Vogeley, K. (2003). Self-representation: Searching for a neural signature of self-consciousness. *Consciousness & Cognition*, *12* (4), 529-543. 10.1016/S1053-8100(03)00080-1

——— (2011). Den anderen verstehen. *Spektrum der Wissenschaft*, *8*

Norman, W. T. & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, *4* (6), 681-691. 10.1037/h0024002

Oyserman, D., Sakamoto, I. & Lauffer, A. (1998). Cultural accommodation: Hybridity and the framing of social obligation. *Journal of Personality Psychology*, *74* (6), 1606-1618. 10.1037/0022-3514.74.6.1606

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14* (1), 30-80. 10.1016/j.concog.2004.10.004

Pegna, A. J., Khateb, A., Michel, C. M. & Landis, T. (2004). Visual recognition of faces, objects, and words using degraded stimuli: Where and when it occurs. *Human Brain Mapping*, *22* (4), 300-311. 10.1002/hbm.20039

Phillips, W., Baron-Cohen, S. & Rutter, M. (1992). The role of eye contact in goal detection. Evidence from normal infants and children with autism or mental handicap. *Development and Psychopathology*, *4* (3), 375-383. 10.1017/S0954579400000845

Reddy, V. (2008). *How infants know minds.* Cambridge, MA: Harvard University Press.

Roberts, T. & Bruce, V. (1988). Feature saliency in judging the sex and familiarity of faces. *Perception*, *17* (4), 829-840. 10.1068/p170475

Sander, D., Grandjean, D., Kaiser, S., Wehrle, T. & Scherer, K. R. (2006). Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion. *European Journal of Cognitive Psychology*, *19* (3), 470-480. 10.1080/09541440600757426

Schank, R. C. & Abelson, R. P. (1977). *Scripts, plans, oals and understanding. An inquiry into human knowledge structures.* New York, NY: Erlbaum.

Schweinberger, S. R., Huddy, V. & Burton, A. M. (2004). N250r: A face-selective brain response to stimulus repetitions. *NeuroReport*, *15* (9), 1501-1505. 10.1097/01.wnr.0000131675.00319.42

Schyns, P. G., Bonnar, L. & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13* (5), 402-409. 10.1111/1467-9280.00472

Secord, P. F., Dukes, W. F. & Bevan, W. (1954). Personalities in faces: I. An experiment in social perceiving. *Genetic Psychology Monographs*, *49* (2), 231-279.

Stephens, G. L. & Graham, G. (2000). *When self-consciousness breaks: Alien voices and inserted thoughts.* Cambridge, MA: MIT Press.

Sui, J. & Han, S. (2007). Self-construal priming modulates neural substrates of self-awareness. *Psychological Science*, *18* (10), 861-866. 10.1111/j.1467-9280.2007.01992.x

Tomasello, M. (1999). *The cultural origins of human cognition.* Cambridge, MA: Harvard University Press.

Tranel, D. & Damasio, A. R. (1985). Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics. *Science*, *228* (4706), 1453-1454. 10.1126/science.4012303

Uleman, J. S., Blader, S. L. & Todorov, A. (2005). Implicit impressions. In R. R. Hassin, J. S. Uleman & J. A. Bargh (Eds.) *The New Unconscious* (pp. 362-392). New York, NY: Oxford University Press.

van den Stock, J., Righart, R. & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, *7* (3), 487-494. 10.1037/1528-3542.7.3.487

Vecera, S. P. & Johnson, M. H. (1995). Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Visual Cognition*, *2* (1), 59-87. 10.1080/13506289508401722

Vetter, P. & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness & Cognition*, *27*, 62-75. 10.1111/j.1468-0017.2006.00298.x.

Vogeley, K. (2012). *Anders sein. Hochfunktionaler Autismus im Erwachsenenalter.* Weinheim, GER: Beltz.

Vogeley, K. & Newen, A. (2002). Mirror neurons and the self construct. In M. Stamenov & V. Gallese (Eds.) *Mirror neurons and the evolution of brain and language* (pp. 135-150). Amsterdam, NL: Benjamins.

Vogeley, K. & Roepstorff, A. (2009). Contextualising culture and social cognition. *Trends in Cognitive Sciences*, *13*, 511-516.

Volz, K. G. (2008). Ene mene mu insider und outsider. In R. Schubotz (Ed.) *Other minds. Die Gedanken und Gefühle anderer* (pp. 19-30). Paderborn, GER: Mentis.

Vosgerau, G. & Newen, A. (2007). Thoughts, motor actions and the self. *Mind and Language*, *22* (1), 22-43. 10.1111/j.1468-0017.2006.00298.x.

Vuilleumier, P. & Pourtois, G. (2007). Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia*, *45* (1), 174-194. 10.1016/j.neuropsychologia.2006.06.003

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V. & Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron*, *40* (3), 655-664. 10.1016/S0896-6273(03)00679-2.

Williams, L. M., Palmer, D., Liddell, B. J., Song, L. & Gordon, E. (2006). The "when" and "where" of perceiving signals of threat versus non-threat. *NeuroImage*, *31* (1), 458-467. 10.1016/j.neuroimage.2005.12.009

Willis, J. & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17* (7), 592-598. 10.1111/j.1467-9280.2006.01750.x

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13* (1), 103-128. 10.1016/0010-0277(83)90004-5

Yamaguchi, S., Kuhlman, D. M. & Sugimori, S. (1995). Personality correlates of allocentric tendencies in individualist and collectivist cultures. *Journal of Cross-Cultural Psychology*, *26* (6), 658-672. 10.1177/0022022195026006009.

Multiplicity Needs Coherence – Towards a Unifying Framework for Social Understanding

A Commentary on Albert Newen

Lisa Quadt

In this commentary, I focus on Albert Newen's multiplicity view (MV) and aim to provide an alternative framework in which it can be embedded. Newen claims that social understanding draws on at least four different epistemic mechanisms, thus rejecting the idea that there is a default mechanism for social cognition. I claim that MV runs the risk of combining elements that have been described in metaphysically incompatible theories. I will argue that multiplicity needs coherence, which can be achieved by applying the theoretical framework of first-, second-, and third-order embodiment (1-3E; Metzinger 2014) to the study of social cognition. The modified version of this theory, 1-3sE (first-, second-, and third-order *social* embodiment), can serve as a unifying framework for a pluralistic account of social understanding.

Keywords
Direct perception | Embodiment | Interaction | Interactive turn | Mirror neurons | Multiplicity view | Phenomenology | Social cognition | Social understanding

Commentator

Lisa Quadt
lisquadt@students.uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Target Author

Albert Newen
albert.newen@rub.de
Ruhr-Universität Bochum
Bochum, Germany

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Introduction

The multiplicity view (MV) is part of Newen's person model theory (PMT) and claims that individuals apply multiple epistemic strategies to make sense of other people, namely simulation, theoretical inference, direct perception (DP) and primary interaction.[1] He thus interestingly argues against the view that there is something like a default strategy of social understanding. In the following, I will scrutinize MV and, in doing so, attempt to reach three goals: First, I reconstruct the main claims of MV and suggest that the development of such a pluralistic account of social cognition can be seen as contributing to the so-called "interactive turn" (Overgaard & Michael 2013; section "The multiplicity view"). MV has the potential

1 For a brief explanation of the terms, see Newen this collection, pp. 1-2.

to integrate bodily and interactive contexts, while also paying more attention to the phenomenology of social encounters. Second, I argue that current pluralistic depictions of social cognition – of which MV is a clear example – run the risk of operating under (often implicit) contradictory background assumptions. In the section "Multiplicity needs coherence", I first show how and why different social cognitive mechanisms have been described under different sets of metaphysical assumptions. Since these assumptions are often contradictory, a coherent version of MV cannot simply claim to combine them. I then go on to argue that the concept of DP as an epistemic mechanism is either metaphysically incompatible with simulation and theorizing, empirically implausible, or – if it is re-formulated so that it fits a representationalist description – does not meet the goal of integrating embodiment and phenomenology anymore. I will thus claim that DP should be used as a phenomenological rather than epistemological concept. My third goal is then to suggest novel ways of adopting a pluralistic perspective on social cognition, while remaining in metaphysically coherent territory. Metzinger's theory of first-, second-, and third-order embodiment (1-3E) is a conceptual framework that combines representationalist and non-representationalist levels of analysis in order to show how a specific phenomenal quality (e.g., phenomenal selfhood) can arise within an embodied system (Metzinger 2014). Metzinger claims that phenomenal properties are computationally grounded in a representation of one's body (the "body model", ibid., p. 273), which in turn is physically implemented by bodily and neural structures. I aim to apply this idea to the study of social understanding (section "1-3sE – Levels of social embodiment"). This application enables a more fine-grained depiction of different phenomenal qualities in social encounters and shows their putative relation to representational and physical counterparts. I ask which parts of the body model could potentially be shared and thus be exploited for a skillful navigation of an individual's social environment. In a last step, I sketch the physical grounds of social cognition.

2 The multiplicity view

The multiplicity view (MV) is part of Albert Newen's person model theory (PMT), which provides a rich and detailed account of social understanding. It attempts to answer two central questions in the research field of social cognition, which the author neatly differentiates and then again integrates into a comprehensive theory. The first question asks which epistemic strategy humans use to access the mental states of others and to gather information about them. Approaches advocating Simulation Theory (ST; e.g., Goldman 2006), as well as direct perception (DP; e.g., Gallagher 2008), have attempted to yield an answer to that question, while Theory Theory (TT; e.g., Gopnik & Meltzoff 1997) and Narrative Practice Hypothesis (NPH; e.g., Hutto 2008) focused on a second question: How is the information we obtain to understand others stored and organized? By sorting out these questions, Newen shows that different theories have tried to tackle different problems, which I believe to be a very useful and fruitful contribution to the research field. It reveals that the four main theories mentioned above are less competitive than originally thought, since, on closer examination, they actually aim to give answers to different questions. This viewpoint enables one of Newen's main arguments, namely that each of these approaches can be merged into one unified account of social understanding. He takes three steps in arguing for his theory. In a first step, he differentiates between the two questions in the research field of social cognition mentioned above, thus setting up a dividing line between the vast manifold of different approaches and theories. Secondly, the author puts forth a pluralistic account of social cognition, the multiplicity view (MV). In doing so, he attempts to answer the first question discussed earlier. In a third step, Newen tackles the second question of how knowledge about other people is organized and stored. He claims that this happens through the formation of so-called *person models*, hence person model theory (PMT; see Newen this collection).

By laying out MV as a pluralistic account of social cognition, Newen aims to steer the discussion in the research field into a different direction, away from debating whether social understanding is a form of simulation, theoretical inference, DP or interaction. Instead, he argues that all four epistemic strategies are applied, depending on the social context (cf. Newen this collection, p. 7). MV is of particular interest, because it reflects two growing convictions in the research field. First, by paying attention to DP and interaction, it does justice to demands that arose in the so-called "interactive turn" (Gallotti & Frith 2013; Overgaard & Michael 2013) and can thus be seen as part of the movement itself. The interactive turn claims that researchers have not paid enough attention to the phenomenology of social encounters (Gallagher 2001), the interactive contexts in which most social situations are embedded (De Jaegher & Paolo 2007) and the role of the body and emotions in social cognition (Schilbach et al. 2013). This directly relates to MV, since it aims to include intuitive ways of social understanding that do not necessarily require simulation and theoretical inference and thus to widen the theoretical scope towards less "cognitivist" views. The second conviction is that there is more to social cognition than a single all-purpose mechanism (Adolphs 2006, p. 30; Fiebich & Coltheart in press).[2] (Human) social cognition obviously is manifold; it has many aspects that are not only phenomenologically distinct (just think of the different experiences you have when trying to figure out your advisor's somewhat cryptic Email, or when trying to make your 4 year-old eat her spinach), but also draws on several cognitive mechanisms that are differently implemented. It therefore makes sense that we can find something useful in each of the four theoretical approaches discussed so far; while ST and TT are plausible accounts to describe and explain "higher-level" social cognition that requires quite sophisticated skills, other theories such as DP or interaction theory cover more intuitive ways of understanding others. Merging them into a comprehensive theory seems to be a natural next step.

Newen claims that

> [t]here is no standard default strategy of understanding others, but in everyday cases of understanding others we rely on a multiplicity of strategies which we vary depending on the context and on our prior experiences (and eventually also triggered by explicit training). (this collection, p. 7)

How does he arrive at this conclusion? Newen argues against the view that only *one* of the mechanisms that have been proposed to be important for social cognition (simulation, theorizing, DP and primary interaction) can plausibly be viewed as the default strategy by which humans understand each other. The main argument against such a single-mechanism view is that their activation seems to be highly context-dependent. Simulation, according to Newen, presupposes similarity between two interacting individuals. Theorizing only applies in complex social situations which need explicit and thoughtful disambiguation. Encountering someone of who we already have rich prior information activates DP, while social situations that are easy to understand can be disambiguated by primary interaction. Thus, Newen concludes that "[o]nly the combination of all four strategies, in full sensitivity to the context and applied on the basis of our experience in successfully using the strategies, makes us experts in understanding others" (ibid., p. 7).

3 Multiplicity needs coherence

While this surely is an attractive way to describe social understanding, and does justice to its oft-proclaimed manifoldness, these mechanisms have been described in several theoretical frameworks that operate under different (and partly contradictory) metaphysical

2 Such a view can already be found in Goldman's work. He endorses a hybrid account of mindreading, which describes "a number of ways to blend simulation and theorizing elements into a mosaic of mindreading possibilities" (Goldman 2006, p. 43).

background assumptions.[3] Thus, a simple combination of them does not come easily. Simulation and theory-based inference have been described within a computationalist, cognitivist framework which often assumes that the mind is mainly a representational and internal device (Bruin & Kästner 2012), i.e., a functional structure locally realized in the brains of individual organisms. Bodily and environmental structures play at most an enabling or causal role for a specific *internal* mechanism. In contrast, DP and primary interaction, both of which are concepts stemming from the phenomenological tradition, have their roots in an enactive account of cognition (cf. Gallagher 2008, p. 537), thus rejecting basic metaphysical assumptions of cognitivism (e.g., representationalism, reductionism, mechanistic explanations; Rowlands 2009).[4] The theoretical background of DP and primary interaction views the mind as a non-representational, relational device which emerges within the skillful interaction between organism and environment:

> The enactive interpretation is not simply a reinterpretation of what happens extraneurally, out in the intersubjective world of action where we anticipate and respond to social affordances. More than this, it suggests a different way of conceiving brain function, specifically in nonrepresentational, integrative and dynamical terms. (Gallagher et al. 2013, p. 422)

More specifically, enactive and phenomenological approaches to social cognition not only see the body as part of cognitive processing, they also assign a very important status to interaction. While enactive theories display interaction as (at least possibly) *constituting* social cognitive processes (De Jaegher & Paolo 2007, p. 493), traditional mindreading theories have not even considered interaction to be an element which could influence social cognition (cf. Fuchs & Jaegher 2009, p. 466).

There are several reasons why ST and TT have been spelled out in a more cognitivist set of assumptions, while DP and primary interaction have been described in reference to an enactive framework. Although their roots in the history of ideas plays an important role, there are deeper systematic reasons why it makes sense to couch them in different sets of metaphysical assumptions. To see this, consider the relation between the external world and internal processing in either framework. A rather cognitivist view assumes that the task of the brain is to figure out the outside world and that this world is *internally represented*.[5] Since other people belong to this world outside of one's own mind, it follows that the causes for their behavior need to be inferred by internal representation processing as well. Because it is assumed that the brain is the only mental organ (Hohwy this collection), the *location* of (social) cognitive processing thus can be said to be inside one individual's head. Simulation and theorizing fit neatly into this picture of the mind; they are inference processes which function to disambiguate social input and are implemented by specific neural mechanisms. By contrast, an enactive view of social cognition as has been described by De Jaegher and colleagues and advocated by Gallagher, presupposes two different assumptions. First, in order to assume that interaction dynamics carry as much of the "cognitive load" to understand other minds as is proposed, a relational view of the mind enters the picture. It is important to understand that an enactive view is not the same as an externalist view, which could be compatible with assumptions of

3 I am well aware of the fact that there are many shades of both cognitivist and enactive views. I will therefore focus on the views of the authors that have been cited by Newen in the target paper. For a general introduction, see for example Thompson (2010); Varela et al. (1993); Rowlands (2009).

4 The difference between enactive and phenomenological theories seems to boil down to the explanatory scope. While enactivism explicitly claims to offer a radically different alternative to cognitivism and thus builds a proper account of cognition (Varela et al. 1993) phenomenology is mostly seen as a description of experiential phenomena (Gallagher 2008).

5 Although this seems to be a rather "old" view, it is currently celebrating a comeback. Jakob Hohwy, for example, claims that the consequences of advocating predictive processing (2013; see also Clark 2013a) are to adopt a fully internalist picture of the mind. In his words, there is an "evidentiary boundary" (Hohwy 2014, p. 6) between what has to be inferred (viz., hidden causes in the external world) and the inference device (the brain). Accordingly, all the processing takes place within this boundary, which happens to be the skull (cf. ibid., p. 8). Please note, though, that both Clark and Seth propose a more embodied perspective on prediction (Clark this collection; Seth this collection).

the cognitivist camp (cf. Rowlands 2009, p. 54). The mind is, according to such an enactive perspective, neither internal nor external; it constitutes itself within the relation (hence *relational*) between an embodied agent and its environment (cf. Di Paolo & Thompson 2014, p. 68; Engel et al. 2013, p. 202). Such a view enables the claim that interactions are examples of this unfolding mental process and thus constitute social cognition. This claim is incompatible with an internalist perspective, which does not ascribe any constitutional power to mind-external properties.

Furthermore, if the external world and the minds of others could be *directly* perceived without further mental processing or inference, neither simulation nor theoretical inference would be needed. This is exactly the point of the non-cognitivist camp, as becomes obvious in this quote by Newen: "The mental states of others are not hidden, and need not to be inferred on the basis of perceiving the behavior; rather, behavior is an expression of the mental phenomena that, in seeing the behavior, is also directly seen" (this collection, p. 5). What does it mean that something can be *directly* seen? Gibson (1979) introduced DP in relation to his famous conception of "affordances": "The affordances of things for an observer are specified in stimulus information. They seem to be perceived directly because they are perceived directly" (Gibson 1977, p. 79). Importantly, the direct perception of affordances is possible because, according to Gibson, affordances are physically *real* (i.e., they exist independent of the perceiving subject) and as such are perceivable properties of objects in the environment (cf. 1979, p. 129). Note how this is crucially different from a view which assumes that object properties need to be mentally represented, thus requiring an intermediary step.[6] However, Gallagher makes explicit in a footnote (cf. 2008, p. 537) that his conception of DP is not to be *entirely* equated with a Gibsonian notion of the term. Gallagher emphasizes that he does not deny the underlying complexity of perceptual processing, much rather he counts those processes as belonging to perception. He thus puts forth the conception of "smart perception":

> But this informing process is already built into the perceptual process so that as I consciously perceive, my perception is already informed by the relevant sub-personal processing. I don't first perceive and then add memory in order to recognize my car. My perception, in this sense, is direct even if the sub-personal sensory processing that underpins it follows a complex and dynamic route. (ibid., p. 537)

Even with that kind of definition, his view still presupposes that there are properties of external objects that can be "directly" picked up, that exist independently from the perceiving subject. As such, it is indeed *reminiscent* of a Gibsonian conception. The difference between cognitivist and non-cognitivist pictures of social cognition, in the cases that I just described, seems to boil down to the metaphysical assumption of whether or not there are hidden causes in the outside world that require an inference or representational mechanism in order to access and process them. While ST and TT clearly assume such a view, DP denies it. Therefore, I claim that MV cannot simply combine theoretical elements that draw on such considerable metaphysical differences.

Another important difference between these theoretical approaches is how each treats the issue of phenomenology. While the experiential nature of social encounters plays at most a minor role in mindreading theories, such as ST and TT, the phenomenal level is of paramount importance for the enactive camp, who advocate for DP. This becomes most obvious in the claim that the experienced smoothness and immediacy of social interactions tells us something about the epistemic access to other minds. However, "directness" as a concept in academic research is relative to a specific level of description. Let me explain this in more detail. Consider Gallagher's argument that smart perception is a subpersonally informed mechanism (cf.

6 In the following, I will use the requirement of intermediary steps as the distinctive feature that differentiates directness and indirectness. In doing so, I follow De Vignemont: "There is a direct access if and only if the causal transmission of information is direct and does not involve intermediary steps" (2010, p. 291).

2008, p. 537) that directly enables an individual to perceive the minds of others without "additional mental effort." It is based largely on the rapid activation of mirror neurons (30-100ms, ibid., p. 541), such that he claims a distinction between a merely perceptual process and an additional mental process does not make sense. In his words:

> A distinction at the neural level between activation of the visual cortex and activation of the pre-motor cortex does not mean that this constitutes a distinction between processes that are purely perceptual and processes that involve something more than perception. (ibid., p. 541)

The question that follows is how one should individuate mental mechanisms, and I suggest that *functional properties* are much more substantial and conceptually relevant individuation criteria than temporal properties. It is, to me, highly questionable whether temporal correlation justifies assuming that there is mechanistic inseparability. The functional role of a mental mechanism seems a much less arbitrary criterion. Furthermore, it enables a more fine-grained view of the subpersonal processes that underlie social cognition. Instead of talking about perception—which could include all processes if only they are activated in a more or less specific amount of time—it is possible to take a closer look at which brain region correlates with which mechanism. If mechanisms are individuated by their *functional role* instead of the temporal properties of the physical realizers of this functional role, it makes sense to assume that the visual system and the mirror neuron system are distinct. If they are, however, it is unfeasible to speak of "smart perception". This concept presupposes that perceptual and post-perceptual processes can coherently be described as *one* mechanism, which I reject. Additionally, the concept of "direct perception" does not apply anymore either, since mirror mechanisms should be seen as a functionally distinct and therefore intermediary step in the process of understanding others. I thus conclude that DP—as described by Gallagher—does not coherently apply to the subpersonal level of description.

This relates to my main point, namely that there are different levels of description at which a phenomenon can be scrutinized. At the phenomenological level, DP can be described as the experience of *directly* and *immediately* perceiving the other person's mental states. I walk into my living room, I see my friend's face and I experience myself as instantaneously knowing that she is really upset. However, this experiential quality of directness is brought forth by a subpersonal process, which is indirect, as I have argued above. At any other level of description, therefore, directness does not apply. In this view, DP is a phenomenal quality of some mental states and should thus not be confused with the epistemic mechanism *itself*. The simultaneity in our everyday experience does not justify anything on other levels of description. I therefore argue that DP should be treated as a phenomenal quality of *some* social encounters instead of assigning it the status of an epistemic strategy to access other minds.

Note that Newen does not explicitly support a phenomenological or enactive view of the mind, nor does he make any claims about the metaphysics of social cognition. What he does do, however, is emphasize Gallagher's conception of DP and primary interaction as being the main sources for an epistemic access to other minds (cf. Newen this collection, p. 8). If Newen was to reject the strong claims of a non-representational view of (social) cognition, however, it is questionable how closely his notions of DP and primary interaction, as core concepts of his theory, actually relate to their original formulations. This leaves us with two options. The first is to assume that Newen fully endorses the views of his oft-cited colleague. In this case, the problem of compatibility becomes obvious. The second, and more likely possibility is that the author does not support DP and primary interaction with all their metaphysical implications. It indeed seems that he rather re-formulates both concepts so that they possibly fit into a representational framework. According to Newen (this collection, p.5), DP is realized by a process of pattern recognition and primary in-

teraction – although Newen explicitly cites Gallagher & Hutto (2008) – is characterized as follows: "[...] I notice a social act being directed towards me and so start to interact, such that a standard interaction is realized, which may be nonlinguistic but may also involve linguistic communication [...]" (Newen this collection, p. 7). What is problematic here is that one of the most interesting and valuable features of MV gets lost, namely its potential to fulfill demands of the interactive turn. A true fulfillment would require widening the theoretical scope of social cognition by going beyond the study of individual brains and considering bodily, interactive and phenomenological processes more carefully.

What needs to be reconciled and made conceptually consistent is thus our choice of a specific, unified methodological framework—our overarching theoretical approach of simulation, theory-based inference, DP and primary interaction—since they all describe important aspects of social understanding. It should be a common aim to work with a coherent set of metaphysical assumptions, since whether or not one agrees on either set of background assumptions has important implications for both theoretical and empirical research. Not only does that decision influence our choice of the *unit of analysis,* i.e., how we frame the explanatory unit for empirical research. For a long time, this unit has been one individual observing another. It has been claimed, however, that this does not properly reflect the real nature of social cognition, and thus a shift is needed:

> The explanatory unit of social interaction is not the brain, or even two (or more) brains, but a dynamic relation between organisms, which include brains, but also their own structural features that enable specific perception-action loops involving social and physical environments, which in turn affect statistical regularities that shape the structure of the nervous system [...]. (Gallagher et al. 2013, p. 422)

When an enactive or phenomenological perspective is adopted and the status of interaction as constituting social cognition is accepted, this adds an additional *level of analysis* (i.e., an "interactionist stance"; De Jaegher et al. 2010) while erasing one that is profound and fundamental for most researchers: representation. Furthermore, the shared goal to pay more attention to the body, interaction and phenomenology comes with many methodological challenges. For all these reasons it should be in the common interest of the research field to find a way to ease the tensions.

As I have shown, Newen tries to combine four elements that might not be entirely compatible. However, the core of his idea is highly valuable, and certainly should not be rejected. What his pluralistic account of social cognition claims is that there are low-level social mechanisms that mainly rely on interaction and do not need complex or explicit thought, while higher-level, sophisticated mechanisms play a just as important role for the phenomenon. While some social situations require processes that allow complex thinking, other contexts can be intuitively disambiguated. In what follows, I will sketch an alternative framework, based on Metzinger's theory of three-level embodiment, which I claim is able to integrate the four elements while operating on coherent background assumptions. Additionally, it has the potential to fulfill the demands of the interactive turn by paying more attention to interactive contexts, the role of the body and the importance of phenomenology.

4 1-3E – First-order embodiment, second-order embodiment, third-order embodiment

Before I describe how the framework of 1-3E itself can be exploited for a pluralistic picture of social cognition, let me describe the framework in more detail. Metzinger's goal is to provide a framework which shows how the experience of being a self is generated within an embodied system (cf. Metzinger 2014, p. 272). The basic assumption is that experiential phenomena (such as phenomenal selfhood) can be described at several different levels: they have a specific phenomenal quality (i.e., phenomenological level of description), which is brought forth by under-

lying computations and representations (i.e., computational/representational level of description). These are implemented by their physical counterparts (i.e., implementational level of description). 1-3E is a theory about the grounding relations between them, that is, the grounding relations holding between phenomenal properties of representational states and their physical and computational resources. In a broader context, Metzinger claims that "the self" is not a thing or an entity (2004), but rather the phenomenal product of a complex computational process which happens to take place in embodied systems. If that is the case, however, the following question arises: How exactly is the experience of being a self generated within an embodied system? In other words, what are the grounding relations of phenomenal selfhood?[7]

Metzinger introduces three levels: first-order embodiment, second-order embodiment and third-order embodiment (Metzinger 2006, 2014). Importantly, these concepts not only describe different levels of embodiment and their relation *within* one system, they also refer to different *classes of systems* which possess different kinds of embodiment. To see this, think of the following three systems which all possess a body and some sort of skillful behavior: a worm, an advanced robot (e.g., the "starfish", see Metzinger 2007), and a human in a waking state. As for the worm, it is safe to say that, in order to navigate its environment, it directly exploits its physical (i.e., bodily) resources. It is highly unlikely, however, that one would find any rule-based computation over an explicit symbol-like representational structure in the worm's nervous system. In Metzinger's terms, this kind of system possesses first-order embodiment (1E system). In contrast to this rather rudimentary kind of embodiment, 2E systems (i.e., systems which possess second-order embodiment) do unconsciously represent themselves *as* embodied. This means that they have some kind of body model that can be exploited by the system in several ways (e.g., as a functional tool for motor control) and sustains skillful interaction with the environment. Importantly, 2E enables counterfactual representation, i.e., the ability to represent possible states without actual execution. The body model thus functionally underlies both physical and virtual behavior (see Cruse & Schilling this collection). What 2E systems are lacking, however, is a *phenomenal* representation of themselves *as* embodied systems. While a robot like the starfish can *use* its unconscious body representation to steer movements, it does not *experience* itself as doing so. Only systems that possess third-order embodiment (3E systems) experience this phenomenal quality of being an agent that owns a body. Humans in non-pathological waking states, for example, possess this kind of embodiment. Along with the ability to use their body model in the same way as 2E systems do, they have the additional sense of owning and controlling this model (cf. Metzinger 2014, pp. 274–275). Interestingly, it is also here that we once again find the phenomenology of "directness" and "immediacy". It is important to note that 2E and 3E systems always possess lower levels of embodiment as well, since they build onto each other and higher levels presuppose the existence of lower levels. In this way, 1-3E can be seen as a grounding theory. To briefly summarize, systems that phenomenally represent themselves as embodied agents possess 3E. Phenomenal properties of states, described at this level, are computationally grounded by referring to a unified representation of the body – second-order embodiment. This unconscious body model, in turn, is grounded in physical and bodily resources, which are described at the lowest level of the hierarchy.[8]

Metzinger is clear about the relation between 2E and 3E; the representational content

7 "It is the problem of describing the abstract computational principles as well as the implementational mechanics by which a system's phenomenal self-model (PSM; cf. Metzinger 2003, 2007) is anchored in low-level physical dynamics, in a maximally parsimonious way, and without assuming a single, central module for global self-representation." (Metzinger 2014, p. 272)

8 I have argued before that a simple combination of cognitivist, representational, and enactive, non-representational perspectives results in a metaphysically incoherent view. One could ask why it should now be possible for 1-3E to put together non-representational and representational levels of description. As I have described earlier, most enactive theories reject representations entirely (e.g., Fuchs & Jaegher 2009, p. 466). That is one important reason why such a view is incompatible with representational theories. Grounding theories, however, take a different perspective on representations. They view them as *grounded* in bodily processes (cf. Pezzulo et al. 2013, pp. 6). As such, representations can be seen as a phenomenon that gradually emerges within an embodied system (cf. Metzinger 2014, p. 278).

of 2E is "elevated to the level of global availability and integrated with a single spatial situation model plus a virtual window of presence" (2014, p. 274). However, one thing that remains relatively vague in his theory is the relation between 1E and 2E. The problem I see here is that Metzinger does not explicitly describe what actually grounds 2E and which role bodily structures play besides that of yielding a grounding relation.[9] A 1E system is defined as a "purely physical, reactive system", which adapts to its environment by exploiting its physical resources. This is not, in my view, what is being represented by a 2E system, which represents itself "as an embodied agent" (ibid., p. 273). What is needed is a more detailed and specific description of 1E and its relation to 2E. Therefore, the discussion of 1E in my own proposal is twofold. First, I analyze the low-level mechanisms that can be described at this level, claiming that they enable basic social skills (e.g., coupling). Second, I describe which neural, bodily and perhaps even extra-bodily structures most likely underlie social processes that are located at the level of 2sE.

There is one important aspect of 3E that I wish to describe in more detail as it will be crucial for my theory. Metzinger distinguishes two kinds of phenomenal properties instantiated by conscious representational states; they can be either *transparent* or *opaque*. Notice that he uses those terms in a rather counterintuitive way I will try to make sense of in the following.[10] An analogy that might help to do so is to think of the difference between a freshly cleaned and a quite dirty window front. In the first case, when the glass is transparent, we can see everything behind it while not perceiving the glass *as* a medium we are looking through. However, if the glass is dirty and opaque, we will not only have trouble seeing the things behind it, we will also perceive the window *itself as* something we are looking through.[11] In analogy, consider mental states (and their processing stages) as either transparent or opaque. A mental state is opaque when it is experienced *as* a representational state. A quite straightforward example is explicit thought where an individual is consciously aware of the fact that she is thinking. The process of representation *is represented as such* in this case, and is therefore opaque. In contrast, if a state is transparent, earlier processing stages are not phenomenally represented; they are not part of the experience of an individual. In the case of phenomenal selfhood, for example, all that is experienced is the sense of being a self in a world. The fact that this experience is a representational process is not part of its phenomenal content. Note that the distinction between phenomenal properties of epistemic mechanisms (such as computations and representations) and epistemic mechanisms themselves is central to the concept of transparency. If we do not experience that a specific phenomenal state is generated subpersonally, when the underlying processes are not elevated to the level of experience, all we experience is the subjective, phenomenological profile of that state. Such a claim is only valid, however, if we assume that these two levels are actually distinct, which seems to be denied by some philosophers in the phenomenological tradition.

In what follows, I will modify parts of the 1-3E framework in order to make it suitable for a pluralistic view of social understanding. The basic scaffold of the theory is retained, since its hierarchical structure is helpful for describing a multi-facetted phenomenon like social cognition. It also offers the possibility for future research to pair 1-3E and 1-3sE with other hierarchical theories of cognition, such as the predictive processing framework (PP; Clark 2013b; Hohwy 2013). PP has not only been described as a very promising theory to unify perception, action and cognition (Clark 2013b), it has also been fruitfully applied to social cognition (Kilner et al. 2007). 1-3sE has the potential to integrate this explanatorily powerful approach, the details of which can be spelled out in future research, but cannot be pursued in this commentary. I furthermore adopt the idea that different levels

9 He gives, though, an example of phenomenal dream states, showing how (parts of) the body model is grounded in bodily structures and processes. Physical eye movements, in this case, most likely ground the phenomenal experience in lucid dreaming (cf. Metzinger 2014, p. 276).

10 For a more detailed description of former usage of the terms, see Metzinger 2003, pp. 345–358.

11 Metzinger uses a similar example: "With regard to the phenomenology of visual experience transparency means that we are not able to see something, because it is transparent. We don't see the window, but only the bird flying by." (2003, p. 358)

of embodiment represent different levels of sophistication and complexity in a system. In order to strengthen this idea and to give an even more differentiated view of social understanding, I aim to make the difference between transparent and opaque social states more obvious. While the general distinction between transparency and opacity is retained, I will modify this aspect in order to make it fruitful for social understanding. To do so, I introduce the concept of "3sE+", which describes experiences in social situations that need explicit and conscious thinking.

Transparency makes it furthermore possible, according to Metzinger, to distinguish one's own body from that of others (cf. Metzinger 2014, p. 274). However, there is an objection I wish to make about this point. I claim that a self-other distinction that functionally serves to identify one's own body in contrast to those of others is already present at the level of 2sE and thus can be achieved without *phenomenally* representing one's body. I will argue for this claim in more detail in the next section.

Additionally, my proposal offers novel ways to enrich Metzinger's original account. He claims that the functional structure of the body model opens a window into social cognition (cf. ibid., p. 273). However, I suggest that this could be a bidirectional relation. There are hints in the literature that being immersed in a social environment is crucial and formative for more general cognitive skills and their development. For example, anecdotal evidence shows that emotional neglect of caregivers severely impairs the physical and mental development of children (Zimmer 1989). Empirical research furthermore shows that the presence, interaction, perception and emotional engagement of and with others shape self-related body representations (e.g., Furlanetto et al. 2013; Schilbach et al. 2013). Longo & Tsakiris (2013) thus conclude that this line of research suggests a strong connection between first-person and so-called second-person (Schilbach et al. 2013) processes, which needs to be considered by researchers of each camp: "Such findings support a model of first-person perspective according to which our sense of self is plastically affected by multisensory information as it becomes available during self-other interactions" (Longo & Tsakiris 2013, p. 430). I thus conclude that it should not only be considered how the development of a self-model influences social cognition, but also which role social processes play in forming such a self-model. This opens interesting and new questions for research on both social cognition and the self. One could ask, for example, whether some social cognitive skills are necessary for the development of a stable self-model or whether there are "genuinely social" parts of the self-model.

5 1-3sE– Levels of social embodiment

In this section, I will introduce an alternative framework in which I describe different processing stages of social understanding as different levels of social embodiment. Before I go into detail about how to apply 1-3E to social understanding, let me motivate my strategy here. I have already pointed out why MV yields an attractive theoretical assumption for research on social cognition. It allows, to briefly repeat, the integration of different aspects of a manifold phenomenon and thus aims to give a comprehensive perspective that is able to encompass sub-areas of interest and research. The advantage of couching MV into 1-3E is that its hierarchical nature affords this integration at different levels of description, while operating on a set of coherent background assumptions. As a grounding theory, it suggests how different levels of analysis relate and at the least has the potential to assign an important role to aspects that lay outside an individual brain. As such it can also do justice to demands from the interactive turn, viz. the consideration of interaction dynamics and their possible role for social cognition as well as taking the phenomenology of social encounters seriously. However, MV suffers from the problem of metaphysical incompatibility. 1-3E, on the other hand, is a representational account that offers a metaphysically sound ground for a manifold phenomenon. My goal is to scaffold a framework for human social cognition, which, as I will argue, can be described as a case of 3E in non-pathological human individuals.

I will now briefly give a rough overview of my proposal of a three-level model of social understanding which I dub "1-3sE" (first-order social embodiment, second-order social embodiment, third-order social embodiment)[12], before I go into detail about what each level amounts to. As in the original version of the framework, levels of social embodiment represent both levels *within* a system and different *kinds* of systems. I thus assume that each social third-order system possesses first- and second-order social embodiment, too. In this commentary, I will focus on describing levels of embodiment within social systems, since this aspect of the framework is of greater importance for a pluralistic view of social cognition.

As previously mentioned, I take it that 1sE fulfills a twofold function: First, it serves as the implementational level of description, showing which physical parts ground higher-level, representational and phenomenal processes. Second, low-level sensorimotor mechanisms subserve basic social interactions (e.g., coupling or synchronization). 2sE involves the instantiation of a model which pre-reflexively represents features of the body. It is assumed that parts of this body model can be shared and thus functionally underlie social cognitive processes that may well operate at the unconscious level, such as imitation, joint attention and action understanding. Finally, 3sE describes cases of consciously experienced social understanding. I claim that there are various kinds of phenomenal experiences in social situations that can be differentiated by applying the concepts of transparency and opacity. Since I consider opaque social mental states to exhibit a very special kind of experience, which is not only rare, but might also entail an additional level of representation, I introduce an extra level: 3sE+. I will now describe the specific levels and their relation in more detail, before I show how my view overcomes the shortcomings of MV.

5.1 Third-order social embodiment (3sE)

Individuals that phenomenally represent themselves as social individuals can be described as social 3E systems (3sE). There are certainly many different ways in which humans experience themselves as being social, but I will focus on those that are mentioned by Newen: DP, personal-level simulation, and explicit theoretical inference.

The concepts of transparency and opacity allow a more fine-grained distinction of different phenomenal experiences of social encounters, as they offer a way to emphasize the similarities and differences between various phenomenal qualities in social situations. DP describes the experience that I can, without being aware of any intermediary steps, understand another person. Importantly, as Zahavi points out, the perceived directness still holds in cases of "unsuccessful" social understanding, such as deception or misunderstandings (cf. 2011, pp. 548–549). Although I can get what you say completely wrong, for example, I would still *experience* myself as *immediately* understanding what you are saying.[13] Since, as I have discussed earlier, the experiential nature of a mental state is not to be equated with its epistemic complexity, we can assume that DP operates on several subpersonal mechanisms. These are, however, not explicitly represented. Hence it makes sense to describe DP as resulting from *transparent* social cognitive states. By doing so, it is possible to keep its phenomenal status as immediate and direct, while not equating this quality with its epistemic status. In contrast, theorizing and personal-level simulation have a quite different phenomenal characteristic. In these cases, the process of *constructing* a specific insight about the other is part of the experience, may this be by explicitly simulating the person (e.g., "If I was her, what would make me excited about having a cat?"), or through theoretical inference (e.g., "People usually own cats to feel less alone, maybe she is excited to have a furry companion now"). They can thus be said to result from *opaque* social cognitive states. What distinguishes transparent from opaque states is the degree to which one's own social cognitive processing, which is directed at the other person, is explicitly represented *as* a process.

12 Note that Schilling and Cruse have already used the abbreviation "1-3SE" to describe levels of situated embodiment. I thus chose a lower case "s" to emphasize the difference (cf. Schilling & Cruse 2008, p. 72).

13 "There is, so to speak, nothing that gets in the way, and it is not as if I am first directed at an intermediary, something different from the state, and then only in a secondary step target it." (Zahavi 2011, p. 548).

However, as already mentioned, I see the need to modify Metzinger's conception of 3E in order to reflect a proper distinction between transparent and opaque social states. I claim that opaque states exhibit an additional level of representation, since the representation process itself is part of the phenomenal experience. In order to emphasize that this is a special and probably rare phenomenon, I introduce the level of "3sE+". Both transparent and opaque social states are certainly to be located at the third level of embodiment, since they possess phenomenal properties. Metzinger suggests that the distinctive feature of 3E in contrast to lower levels is that it enables the system to identify itself with its body (cf. Metzinger 2014, p. 274). The resulting phenomenal properties of self-identification and selfhood stem from the experienced immediacy that comes with transparency (cf. ibid., p. 273). If this is the case, it can be assumed that phenomenal states are not *either* transparent *or* opaque, but that transparency is part of *any* phenomenal state. The degree to which the representation process is explicitly represented varies, transparency and opacity are thus gradually arising properties (cf. Metzinger 2003, p. 358). Additionally, it could well be that there is a constant oscillation between transparency and opacity, depending – for example – on specific contexts and situations. However, opacity and the resulting experiences seem to be more high-level features that can only be found in a small subgroup of species. This is obvious in social understanding, since full-fledged theoretical inference and high-level simulation are not very likely to be found in most non-human animals and human infants. It seems that in the case of opaque states there is an additional level of representation that requires a higher level of sophistication, which should be made more explicit in the hierarchical framework. Transparent and opaque mental states – at least in this case for social understanding – reflect two different kinds of phenomenal experiences that might also have different underlying mechanisms. I thus introduce, in order to do justice to this difference, an additional level of 3sE, namely 3sE+. 3sE+ describes those phenomenal states during which one is aware of the constructing process and which occurs in situations that require this kind of reasoning in order to disambiguate the input. This additional distinction at the level of 3sE enables a more detailed view and underlines the difference between transparency and opacity.

One question that arises at this point is the following. We have assumed that opacity means to phenomenally represent (parts of) the actual process of representation. Does that mean that in the case of theorizing and simulation one would find their underlying representational processes to be subpersonal kinds of theoretical inference and simulation? There are two points that speak against this assumption. First, there are justified worries that the conception of implicit theorizing as an unconscious process stretches the concept of a theory too far (e.g., Blackburn 1992). These arguments against TT have been presented extensively in the literature and I will thus not repeat them here. In the case of simulation, secondly, it seems that subpersonal or low-level simulation does not necessarily generate the phenomenal experience of simulating. Consider the many studies that have been conducted to explore whether the activity of the mirror neuron system can be seen as a kind of implicit simulation that enables social understanding (for a review, see for example Cattaneo & Rizzolatti 2009). In most of these experiments that found mirror neuron activity to be correlated with social understanding, it seems that the phenomenal experience has the character of DP rather than explicit simulation.[14] Such a view, as I hope to have shown, has two advantages. It describes different kinds of phenomenal experiences in social encounters and distinguishes them by referring to the concepts of transparency and opacity.

5.2 Second-order social embodiment (2sE)

Assuming that there is something like a representational body model, we can now ask which

14 Note that this is a speculative claim, since almost none of the studies contain phenomenological reports. It could be fruitful, however, for future research to pay more attention to the experience that participants have in a specific experimental setting. This would help to understand which kind of epistemic mechanism generates which kind of experience.

parts of it can be exploited for social cognition. In order to do so, let me briefly recapitulate how to conceive of this body model. It has been described as a "grounded, predictive body model that continuously filters data in accordance with geometrical, kinematic and dynamic boundary conditions" (Metzinger 2014, p. 273). Furthermore, Metzinger predicts that parts of this model can be shared by individuals: "[…] on a certain level of functional granularity, this type of core representation [i.e., the body model] might also describe the generic, universal geometry which is shared by all members of a biological species" (ibid., p. 273; see also Schilling & Cruse 2012). Together with Gallese he argues elsewhere that the mirror neuron system plays a crucial role in generating a basis for both an "internal model of reality" as well as a "shared action ontology" (Metzinger & Gallese 2003, p. 550). This means, as I take it, that the body model contains information that represents one's own body, but is not completely self-specific. To see this, consider that in order to be shared, representations must not be too specific as to not generalize to the bodies of others. I will come back to this point soon. This consequence worried Newen, leading him to reject the view that mirror neurons form a basis for social cognition:

> Why are mirror neurons not an essential part of understanding others? They represent a type of action or emotion that is independent from a first- or third-person perspective; but the distinction between self and other is an essential part of understanding others (this collection, p. 4).

This raises the question of what exactly it is that can be shared by individuals. Since these considerations are central to the possibility of exploiting the body model for social understanding, I now aim to refute the worry and give a possible answer to the question.

Mirror neurons were discovered in the premotor cortex of macaque monkeys more than 20 years ago. They fire, as is famously known, both when an individual executes and observes an action (Gallese et al. 1996; Rizzolatti et al. 1996; Rizzolatti & Craighero 2004). Although there is considerable controversy about their existence in humans (Hickok 2009), their actual function (Jacob 2008), and their explanatory power (Borg 2007), they are considered by many researchers to form one of the crucial systems for understanding others (e.g., Stanley & Adolphs 2013, p. 512). Mirror neurons are indeed neutral to the agent of an action – they fire whether an action is executed by oneself or another person. Insofar, critics are right to say that it is not obvious how they could provide the important distinction between self and other. However, it seems that there are two important facts left out in this line of thinking. Firstly, it has been suggested that there are inhibition mechanisms that "control" shared representations and provide the basis for a self-other distinction (for a more detailed discussion, see Brass et al. 2009). Secondly, mirror neurons have always been presented as being embedded in a *system* (hence mirror neuron system, e.g., Cattaneo & Rizzolatti 2009; Iacoboni & Dapretto 2006; Rizzolatti & Craighero 2004). This system consists of areas which contain mirror neurons, but also regions which contain neurons that do not have bimodal properties and encode only self-generated actions, as described by Jeannerod & Pacherie (cf. 2004, pp. 131–132).[15] Thus, it is correct that mirror neurons *alone* do not distinguish between self and other. However, this is a rather impoverished view, since they should never be considered in isolation. A similar thought which helps to refute the worry is given by De Vignemont who adopts the view that mirroring can be seen as sharing body repres-

15 "The problem of agent-identification, however, is solved by the fact that other premotor neurons (the canonical neurons) and, presumably many other neuron populations as well, fire only when the monkey performs the action and not when it observes it performed by another agent. This is indeed another critical feature of the shared representations concept: they overlap only partially, and the non-overlapping part of a given representation can be the cue for attributing the action to the self or to the other. The same mechanism operates in humans. Neuroimaging experiments where brain activity was compared during different types of simulated actions (e.g., intending actions and preparing for execution, imagining actions, observing actions performed by other people) revealed, first, that there exists a cortical network common to all conditions, to which the inferior parietal lobule (areas 39 and 40), the ventral premotor area (ventral area 6), and part of SMA contribute; and second, that motor representations for each individual condition are clearly specified by the activation of cortical zones which do not overlap between conditions […]." (Jeannerod & Pacherie 2004, pp. 131–132)

entations (2014a). She argues that shared body representations do not threaten a self-other distinction because they always contain information that is too self-specific to be shared. They are, in her words, "[…] Janus-faced. They face inward as representations of one's body and they face outward as representations of other people's bodies" (De Vignemont 2014b, p. 135).

A closer look at her conception also yields a possible answer to the question of what it is that can be shared with others. De Vignemont argues that it must be a rather coarse-grained representation of one's body, since bodies differ considerably in many aspects like size, gender, posture etc. This representation, which De Vignemont dubs the "body map" (De Vignemont 2014a, p. 289, 2014b, p. 134), contains information about the basic configuration of body parts and thus serves as a functional tool to localize bodily experiences. Irrespective of individual differences of this map, some of its content is so coarse-grained that humans are still able to imitate others or experience vicarious bodily sensations, both of which have been claimed to draw on shared body representations. In other words, what can be shared is that part of the body map whose content is general enough to apply to all kinds of bodies, no matter their differences.

Although this is surely no exhaustive inquiry of the matter, these thoughts provide an idea of how to view 2sE as enabling social cognition: at the representational level, there are parts of the body model which can be shared with others.[16] These parts, however, have to be embedded in a system that also contains self-specific information. Otherwise it would be impossible to attribute an action, an experience or observation to the agent concerned. It now becomes obvious why I claimed earlier that a self-other distinction does not need a *phenomenal* representation of one's body. The unconscious body model and its shared parts seem well furnished to provide such a distinction and thus make unconscious social processes such as mimicry and involuntary imitation possible.

16 Sharing means that representational content overlaps, at least partially. For a more detailed discussion on sharing, see De Vignemont 2014b; Jeannerod & Pacherie 2004.

5.3 First-order social embodiment (1sE)

Although interaction is certainly a topic that has been the least explored by researchers of social cognition, it nevertheless should be considered carefully by any theory that aims to provide a comprehensive view on social understanding. Including interaction is particularly challenging, since most attempts to do so came from proponents of an enactive perspective on the mind. However, I have argued that a pluralistic model of social cognition cannot simply combine enactive claims with cognitive ones (see section 3 "Multiplicity needs coherence"). What is needed is an approach of social understanding that integrates interaction as a phenomenon that most probably does not need explicit, high-level representation. 1sE offers a way to describe such low-level social processes. Knoblich and Sebanz, for example, review several cases of "social coupling". Individuals tend to synchronize their movements if they are sitting next to each other in a rocking chair (cf. Knoblich & Sebanz 2008, p. 2022), a process which can plausibly be described without representation. This sort of "entrainment" (ibid., p. 2023) is a case of coupling during which individuals influence each other's behavior without consciously intending to do so. There are also cases in the animal kingdom that can be described at the level of 1sE, such as the formation and synchronization of fireflies (Suda et al. 2006).

The next step is to depict the implementation of specifically "social parts" of the body model. What physically grounds them is described at the level of 1sE. One buzzword in the research field of social cognitive neuroscience is "the social brain" (e.g., Dunbar 1998; Gazzaniga 1985). This term refers to all the different areas in the brain that have been found to be correlated to cognitive processing in social situations, including, of course, the mirror neuron system. While the investigation of brain regions and their functions for social cognition is a well-established endeavor, it will be more interesting to look at other possibilities of implementing social cognition. The role of interaction for social cognition, for example, has been hotly disputed in the research field. As I have il-

lustrated earlier, some claim that interaction dynamics could *constitute* social cognitive mechanisms (De Jaegher et al. 2010). However, such a view is only sustainable in a radically enactive set of assumptions and as such is not an option for the framework I am suggesting here. What should be considered, though, is whether being in an interaction is *necessary* for some social cognitive states. It has been suggested by recent studies that activation patterns differ depending on the situational context and the degree of emotional engagement in a social situation (Schilbach et al. 2013). These results point to this possibility, but it still needs more careful investigation whether or not they justify the claim that interaction in any way *physically grounds or enables* social cognition.

Such basic and non-representational forms of social understanding have been neglected by the research field for a long time and are in need of more empirical and philosophical investigation. Especially research on joint action and coupled systems is therefore important to sort out 1sE.

6 Conclusion

My first goal in this commentary was to show that MV as a pluralistic view on social understanding is a valuable contribution to the interactive turn. It has the potential to integrate insights from different directions of empirical and theoretical research and thus to yield a comprehensive account on social cognition. However, I argued that such an approach needs careful consideration concerning its metaphysical background assumptions. I demonstrated that parts of MV as laid out by Newen are not fully compatible and that it thus needs a different kind of framework which allows a coherent picture.

I presented an alternative model by applying Metzinger's framework of 1-3E to social cognition, hence 1-3sE. Although the details are still to be spelled out in future research, 1-3sE has several advantages that enable a coherent and fruitful framing of MV. It integrates all four social mechanisms mentioned by Newen and thus can be seen as a pluralistic account of social cognition. What is different, however, is that those four elements are described at different levels of description. As such they all play a specific role in the overall image of social understanding and merge into a manifold, but unified picture. Basic interaction, in this theory, can be accounted for without making radical claims in either direction; we do not need to assume that the mind is relational, as claimed by proponents of the enactive theory. However, we also do not have to ascribe a high level of sophistication to a system in order to be able to interact. In my proposal, interaction (or at least simple interactive mechanisms) can function without any complex representation. Interaction is thus located at the lowest level of the hierarchy, namely 1sE. The next level of social embodiment describes representational and computational processes that subserve social cognition. I showed in which ways a model of one's own body could enable social cognition and which parts of such a model could possibly be shared with others. 2sE encompasses these processes. I further argued that DP should be treated as a phenomenological rather than epistemological concept and should thus be described at the level of 3sE. By doing so, I aimed to avoid mixing up different levels of description and to yield a coherent usage of the term. High-level simulation and theoretical inference have been described at the level of 3sE+, the highest level of the hierarchy, thus doing justice to the fact that they are very special and probably rare cases of social cognition. The application of the notions of transparency and opacity offered a way to emphasize the phenomenological variety that comes with different social situations.

There are still many open questions and this is by no means an exhaustive description of how 1-3sE can be used to frame social understanding. My goal here was to highlight its potential to provide a framework which offers novel ways to (1) incorporate the phenomenal level of description with its representational counterparts, (2) to integrate the role of the body as shaping and grounding social cognitive processes and thus (3) to depict social cognition as a representational, but still embodied ability.

Acknowledgements

I would like to thank the Barbara-Wengeler-Stiftung for their financial support, Thomas Metzinger and Jennifer Windt for giving me the opportunity to be part of this project, two anonymous reviewers for their valuable feedback, and Luke Miller for helpful suggestions on form and content.

References

Adolphs, R. (2006). How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Research*, *1079* (1), 25-35. 10.1016/j.brainres.2005.12.127

Blackburn, S. (1992). Theory, observation and drama. *Mind & Language*, *7* (1-2), 187-230. 10.1111/j.1468-0017.1992.tb00204.x

Borg, E. (2007). If mirror neurons are the answer, what was the question? *Journal of Consciousness Studies*, *14* (8), 5-19.

Brass, M., Ruby, P. & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364* (1528), 2359-2367.

Bruin, L. C. de & Kästner, L. (2012). Dynamic embodied cognition. *Phenomenology and the Cognitive Sciences*, *11* (4), 541-563. 10.1007/s11097-011-9223-1

Cattaneo, L. & Rizzolatti, G. (2009). The mirror neuron system. *Archives of neurology*, *66* (5), 557-560.

Clark, A. (2013a). Dreaming the whole cat: Generative models, predictive processing, and the enactivist conception of perceptual experience. *Mind*, *121* (483), 753-771.

——— (2013b). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181-253. 10.1017/S0140525X12000477

——— (2015). Embodied prediction. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Cruse, H. & Schilling, M. (2015). Mental states as emergent properties - From walking to consciousness. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

De Jaegher, H., Di Paolo, E. & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14* (10), 441-447. 10.1016/j.tics.2010.06.009

De Jaegher, H. & Paolo, E. (2007). Participatory sensemaking. *Phenomenology and the Cognitive Sciences*, *6*, 485-507. 10.1007/s11097-007-9076-9

De Vignemont, F. (2010). Knowing other people's mental states as if they were one's own. In D. Schmicking & S. Gallagher (Eds.) *Handbook of Phenomenology and Cognitive Science* (pp. 283-299). Dordrecht, NL: Springer.

——— (2014a). Shared body representations and the 'whose' system. *Neuropsychologia*, *55*, 128-136. 10.1016/j.neuropsychologia.2013.08.013

——— (2014b). Acting for bodily awareness. In R. Shapiro (Ed.) *The Routledge Handbook of Embodied Cognition* (pp. 287-295). New York, NY: Routledge.

Di Paolo, E. & Thompson, E. (2014). The enactive approach. In R. Shapiro (Ed.) *The Routledge Handbook of Embodied Cognition* (pp. 68-78). New York, NY: Routledge.

Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary Anthropology*, 178-190.

Engel, A. K., Maye, A., Kurthen, M. & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, *17* (5), 202-209.

Fiebich, A. & Coltheart, M. (in press). Varieties of social understanding. *Mind & Language*

Fuchs, T. & Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences*, *8* (4), 465-486.

Furlanetto, T., Cavallo, A., Manera, V., Tversky, B. & Becchio, C. (2013). Through your eyes: Incongruence of gaze and action increases spontaneous perspective taking. *Frontiers in Human Neuroscience*, *7*.

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

——— (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17* (2), 535-543. 10.1016/j.concog.2008.03.003

Gallagher, S., Hutto, D. D., Slaby, J. & Cole, J. (2013). The brain as part of an enactive system. *Behavioral and Brain Sciences*, *36* (4), 421-422. 10.1017/S0140525X12002105

Gallagher, S. & Hutto, D. D. (2008). Understanding others through primary intersubjectivity and narrative practice. In J. Zlatev, C. Shina & E. Itkonen (Eds.) *The Shared Mind: Perspectives on Intersubjectivity* (pp. 1-18). Amsterdam, NL: John Benjamins.

Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 592-609.

Gallotti, M. & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17* (4), 160-165. 10.1016/j.tics.2013.02.002

Gazzaniga, M. S. (1985). *The social brain: Discovering the networks of the mind.* New York, NY: Basic Books.

Gibson, J. (1977). The theory of affordances. In R. Shaw (Ed.) *Perceiving, acting, and knowing. Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ: Erlbaum.

——— (1979). *The ecological approach to visual perception.* Boston, MA: Houghton Mifflin Company.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* New York, NY: Oxford University Press.

Gopnik, A. & Meltzoff, A. N. (1997). *Words, thoughts, and theories. Learning, development, and conceptual change.* Cambridge, MA: MIT Press.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, *21* (7), 1229-1243. 10.1162/jocn.2009.21189

Hohwy, J. (2013). *The predictive mind.* Oxford, UK: Oxford University Press.

——— (2014). The self-evidencing brain. *Noûs*. 10.1111/nous.12062

——— (2015). The neural organ explains the mind. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Hutto, D. D. (2008). The narrative practice hypothesis: Clarifications and implications. *Philosophical Explorations*, *11* (3), 175-192. 10.1080/13869790802245679

Iacoboni, M. & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7* (12), 942-951. 10.1038/nrn2024

Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind Language*, *23* (2), 190-223. 10.1111/j.1468-0017.2007.00337.x

Jeannerod, M. & Pacherie, E. (2004). Agency, simulation and self-identification. *Mind and Language*, *19* (2), 113-146. 10.1111/j.1468-0017.2004.00251.x

Kilner, J. M., Friston, K. J. & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, *8* (3), 159-166.

Knoblich, G. & Sebanz, N. (2008). Evolving intentions for social interaction: From entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363* (1499), 2021-2031. 10.1098/rstb.2008.0006

Longo, M. R. & Tsakiris, M. (2013). Merging second-person and first-person neuroscience. *Behavioral and Brain Sciences*, *36* (04), 429-430.

Metzinger, T. (2003). Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences*, *2*, 353-393.

——— (2004). *Being no one: The self-model theory of subjectivity.* Cambridge, MA: MIT Press.

——— (2006). Different conceptions of embodiment. *Psyche*, *12* (4)

——— (2007). Self models. *Scholarpedia*, *2* (10), 4174.

——— (2014). First-order embodiment, second-order embodiment, third-order embodiment: From spatiotemporal self-location to minimal selfhood. In R. Shapiro (Ed.) *The Routledge Handbook of Embodied Cognition* (pp. 272-286). New York, NY: Routledge.

Metzinger, T. & Gallese, V. (2003). The emergence of a shared action ontology: Building blocks for a theory. *Consciousness and Cognition*, *12* (4), 549-571. 10.1016/S1053-8100(03)00072-2

Newen, A. (2015). Understanding others - The person model theory. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Overgaard, S. & Michael, J. (2013). The interactive turn in social cognition research: A critique. *Philosophical Psychology*, *28* (2), 1-25. 10.1080/09515089.2013.827109

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K. & Spivey, M. J. (2013). Computational grounded cognition: A new alliance between grounded cognition and computational modeling. *Frontiers in Psychology*, *3*.

Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27* (1), 169-192. 10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3* (2), 131-141.

Rowlands, M. (2009). Enactivism and the extended mind. *Topoi*, *28* (1), 53-62. 10.1007/s11245-008-9046-z

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T. & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36* (4), 393-414. 10.1017/S0140525X12000660

Schilling, M. & Cruse, H. (2008). The evolution of cognition: From first order to second order embodiment. In I. Wachsmuth & G. Knoblich (Eds.) *Lecture notes in computer science Lecture notes in artificial intelligence: Vol. 4930. Modeling communication with robots and virtual humans. Second ZiF Research Group International Workshop on Embodied Communication in Humans and Machines, Bielefeld, Germany, April 5 - 8, 2006; revised selected papers* (pp. 77-108). Berlin: Springer.

——— (2012). What's next: Recruitment of a grounded predictive body model for planning a robot's actions. *Frontiers in Psychology*, *3*.

Seth, A. K. (2015). Inference to the best prediction. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Stanley, D. A. & Adolphs, R. (2013). Toward a neural basis for social behavior. *Neuron*, *80* (3), 816-826. 10.1016/j.neuron.2013.10.038

Suda, T., Tschudin, C., Tyrrell, A., Auer, G. & Bettstetter, C. (2006). *Fireflies as role models for synchronization in ad hoc networks. In Proceedings of the 1st international conference on Bio inspired models of network, information and computing systems.* New York, NY: ACM.

Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind.* Cambridge, MA, London: Belknap.

Varela, F. J., Thompson, E. & Rosch, E. (1993). *The embodied mind: Cognitive science and human experience.* Cambridge, MA: MIT Press.

Zahavi, D. (2011). Empathy and direct social perception: A phenomenological proposal. *Review of Philosophy and Psychology*, *2* (3), 541-558. 10.1007/s13164-011-0070-3

Zimmer, D. E. (1989). Wilde Kinder. In D. E. Zimmer (Ed.) *Experimente des Lebens* (pp. 21-47). Zürich: Haffmanns Verlag.

A Multiplicity View for Social Cognition: Defending a Coherent Framework

A Reply to Lisa Quadt

Albert Newen

Lisa Quadt's commentary focuses on my theory about the multiple epistemic strategies humans use to receive information about one other's mental phenomena. She develops a principle worry about the theory's underlying metaphysical foundations, arguing that I am committed to an incoherent metaphysical framework. In this reply, I show that I am not committed to the position she attributes to me and I outline an alternative framework that is my actual background view. I illustrate this framework by discussing emotions and argue that emotions are individuated as integrated patterns of characteristic features. This enables me to combine a representational account of emotions with a theory of direct perception of basic emotions as well as with an understanding of some emotions relying on theory-based inferences. Thus, I have a coherent metaphysics. Finally, I show that the alternative suggested by Quadt has its own problems.

Keywords
Direct perception | Metaphysical foundation | Person model theory | Social cognition | Transparency

Author

Albert Newen
albert.newen@rub.de
Ruhr-Universität Bochum
Bochum, Germany

Commentator

Lisa Quadt
lisquadt@students.uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

1 Introduction

With my PMT (person model theory), I aim to answer two questions. While the first question asks which epistemic strategy humans use to access the mental states of others and to gather information about them, the second question asks how the information we obtain to understand others is stored and organized. The answer to the second question is the core of the PMT. It states that information about other individuals or types of persons is stored and organized in person models and that these are realized on two levels, i.e. the implicit level of person schemata and the explicit level of person images. It further argues that philosophical theories so far have predominately ignored the fact that we usually understand others relying on rich background information concerning them and their situation.

Lisa Quadt's commentary focuses on my theory concerning the epistemic strategies humans use to receive information about others' mental phenomena, and she develops a principle worry about the underlying metaphysical foundations. I am grateful for this challenge, which gives me the opportunity to clarify my background view. The MV (multiplicity view) outlined in the target paper claims that we do not rely on one epistemic strategy alone, as is suggested by most proposals in the literature, but that we rely on a multiplicity of strategies which, for the most part, are implicitly activated on the basis of contextual conditions. These strategies include simulation strategies, theory-based inferences, and direct perception as well as understanding by social interaction and by relying on narratives. Quadt's main worry is that MV may be based on an incoherent metaphysics and is thus inacceptable as it stands. In the first part of her reply she aims to defend the incoherence claim, while in the second part she offers an alternative metaphysical framework. My reply is structured as follows: In the next paragraph I briefly describe how Quadt defends her claim about the supposed incoherence of my metaphysical background and show that I am not committed to the incoherent framework she attributes to me. In the second section, I make explicit my actual background metaphysics (which was not the focus of my article) and argue that it is coherent, reinforcing that I am not committed to the metaphysics that Quadt attributes to my position. Finally, I argue that the alternative metaphysics suggested by Quadt relies on a distinction between transparency and opacity that cannot carry the weight it is supposed to carry.

2 Am I committed to an incoherent metaphysics?

Quadt describes correctly that the MV I advocate combines epistemic strategies that are described in several different positions, including ST (Simulation Theory) (Goldman 2006), TT (Theory-Theory) (e.g., Gopnik & Meltzoff 1997), and IT (Interaction Theory) (Gallagher 2001), as well as theory of direct perception (Gallagher 2008). As a consequence, she presupposes that I am committed to the metaphysical foundations of each of these positions, while each position argues for a distinct epistemic strategy. If I were committed to accepting such metaphysical foundations, I would thereby offer an incoherent metaphysics. Quadt shows this by arguing that Simulation Theory and Theory-Theory, on the one hand, presuppose metaphysical claims that are not consistent with the presuppositions from Theories of DP (direct perception) and ITs, on the other hand (3). Quadt claims that ST and TT are *cognitivist theories* that presuppose internalism, mental representations, and the idea that mental phenomena are private hidden entities to which we have no direct access. To register mental phenomena we have to rely on perceiving the behaviour and expressions of other people and have to *infer* the existence of mental phenomena. Quite the opposite view is taken by the *non-cognitivist theories* of DP and IT. They allow for externalism of mental phenomena (as being realized by two people and their interaction), they deny the existence of mental representations, and they presuppose that mental phenomena are not hidden but directly perceivable. Thus they rely on non-inferential access to mental phenomena by direct perception. The following quote illustrates the main features of the contrast Quadt develops:

> The difference between cognitivist and non-cognitivist pictures of social cognition, in the cases that I just described, seems to boil down to the metaphysical assumption of whether or not there are hidden cause in the outside world that require an inference or representational mechanism in order to access and process them. While ST and TT clearly assume such a view, DP denies it. Therefore, I claim that MV cannot simply combine theoretical elements that draw on such considerable metaphysical differences. (Quadt 2015, p. 5)

My first general reply to this worry is that I only take on the description of an *epistemic strategy* of acquiring and using information

about other people in order to understand them. An epistemic strategy like a simulation (to put oneself in the other person's shoes) or a theory-based inference is not automatically connected to a metaphysical commitment. De facto, the philosophers who are famous for holding ST or TT combine their view with a metaphysical background, but it does not follow that the epistemic strategy they describe *must be combined* with the metaphysical background they offer. We can easily see this for example in the case of two epistemic strategies like theory-based inferences and direct perception of mental phenomena. These can be easily combined in a way that allows that some mental phenomena with intense expressive components like basic emotions (Ekman et al. 1972) can be directly perceived (see below), while complex mental phenomena like propositional attitudes may be at least often inferred if the social understanding cannot rely on honest utterances but only on some ambiguous behavioural cues. Thus, the de facto incompatibility of the metaphysical presuppositions of the two main lines of theories of social understanding does not imply that I am committed to inheriting both presuppositions and that I thus run into an incoherent metaphysics. In fact, I do not presuppose two metaphysical principles for the same mental phenomenon; instead I only need to allow for the application of two epistemic strategies of understanding mental phenomena, which may be applied to different mental phenomena (or to the same type of mental phenomenon in different situations). In the next section I outline my alternative metaphysics and illustrate both that it is coherent and that it can allow for direct perception as one epistemic strategy for registering some mental phenomena.

3 Defending direct perception in an alternative metaphysical framework

In general, I prefer to think of mental phenomena as representational, but I do not see that this prevents me from integrating the epistemic strategy of direct perception. Furthermore, I characterize basic emotions as realized in one individual (individualism but not internalism). At the same time, I remain neutral as to whether joint emotions (e.g. joint enthusiasm about a goal achieved by one's team) have to be analysed as extended emotions. Furthermore, I think that basic emotions are not hidden mental phenomena but can be directly perceived e.g. on the basis of face-based recognition of emotions. Thus, I think that some mental phenomena can be registered non-inferentially. But of course, direct perception of some mental phenomena is *only one* of at least four epistemic strategies that we can use, depending on the context.

To sketch my theory of direct perception I will focus on basic emotions like anger, fear, happiness, sadness, etc. (for a classification of emotions see Zinck & Newen 2008). My metaphysical view of emotional episodes is that they are integrated patterns of characteristic features (Welpinghus & Newen 2012; Newen et al. 2015). Let me use the example of fear as illustrated in Newen et al. (2015): an emotional episode of fear towards an aggressive dog is constituted by the integration of the following characteristic features: (1) a typical physiological arousal that is a consequence of bodily changes due to changes in the autonomic nervous system, including increased heart rate and flat breathing; (2) a typical behavior or behavioral disposition, including flight or freezing behavior; (3) a typical facial expression, gesture, or body posture, etc.; (4) a typical phenomenal experience of fear; (5) a typical (explicit) cognitive evaluation of the dog in front of me (e.g., "This is an aggressive pit bull"). Furthermore, every emotional episode has (6) an intentional object, i.e. the dog in front of me. Features 1–5 are integrated into an (often implicit) appraisal of the intentional object as dangerous. The emotional episode is constituted by the integration of all the characteristic features mentioned so far, including the appraisal. This view allows that in another implementation *some* features would be missing. For example, the explicit cognitive evaluation of the dog as an aggressive pit bull is not necessary to be in fear towards the dog in front of me. Or the facial expression may be inhibited, due to intense training to attain a poker face, yet I may still be in fear. As long as a minimum of features is realized, we still have

an episode of fear. The two main features that are necessary in all emotional episodes are a registration of minimal physiological arousal and an intentional object. The integration of both is needed to have an emotional episode (Barlassina & Newen 2013). But other features may be lacking while still remaining characteristic of most episodes of the relevant type of emotion. One might wonder why I do not include neural correlates. Since I argue from a position of antecedent naturalism, neural correlates are not an extra component in addition to the characteristic features already mentioned above. We might mention neural correlates as an informative aspect for the individuation of certain features of emotion, but we do not have to, since they concern the same features that have already been mentioned, with information accessed in a different manner.

If one accepts the ontology of emotions as individuated by an integrated pattern of characteristic features, it follows that the expression of an emotion by face, body posture, and gestures is a *constitutive* part of the emotional episode (and not a causal consequence). Thus, I do not hold internalism about mental phenomena. Given this theory of the individuation of emotions, I also argue for the thesis that one way of recognizing emotions is by perceiving the relevant pattern (Newen et al. 2015). A recognition of the other person's fear can be attained by directly perceiving the pattern of fear. How can we account for this, while at the same time accepting that the feeling of fear is a private subjective experience in so far as a person still may have the feeling even if she is able to keep a poker face? Perceiving fear is comparable to perceiving a house. Both are processes of pattern recognition on the basis of a minimal package of characteristic features: I can recognize a drawing as one of a house, even if one or two of the characteristic features of a house are missing. How is this possible? Perceiving an object is not a purely passive process, like taking a photograph; it is a constructive process.[1] One important aspect of the constructive process is the enrichment of selected core sensory information. And one way of realizing this enrichment is by the activation of a rich memorized mental image that best suits the core sensory information. If we have learned the relevant pattern of what a house looks like from the outside, and memorized a respective mental image, then seeing a child's drawing initiates an interaction of bottom-up and top-down processes. These include the activation of this stored mental image, such that it enriches the core sensory information to form a perceptual experience of seeing a drawing of a house even if the front door is missing in the drawing.

The same process of pattern recognition takes place in the case of recognizing an emotion like fear. The relevant pattern of fear is formed either on the basis of having personally experienced a situation of fear or on the basis of having observed others in such situations. One thereby acquires a memorized pattern of fear with typical features. If one now observes a person with a typical facial expression in a situation where she is being attacked by a dog, one can see the fear of the person. The perception of fear is realized by seeing the freezing behaviour, the facial expression, and the intentional object (i.e. the aggressive dog), because these features activate as part of the process of perceptual processing the whole pattern of fear. Thus, I can *perceive* fear in the face of the person being attacked by the dog. The theory of perception is one according to which perceptual processing allows for a systematic enrichment of information and for influencing of perceptual processes by memorized images or background knowledge. These top-down influences are discussed under the label cognitive penetration. So I am committed to the view of perception as cognitively penetrated as it is defended in detail in Vetter & Newen (2014). But this does not involve any claims concerning the metaphysical commitments ascribed to me by Quadt in her commentary. Recognition of emotions is analysed in a framework that explicitly allows for mental representations but specifies them in a way that nevertheless allows for direct perception as one form of access to the recognition of

1 All modern theories of perception account for this constructive component, e.g. O'Regan's and Noë's theories of enacted perception (O'Regan & Noë 2001; Noë 2005), as do theories of cognitive penetration (Macpherson 2012; Siegel 2012) and theories of predictive coding (Hohwy 2013; see alsoHohwy this collection; Clark this collection).

emotions. As has been spelled out in detail elsewhere (see Newen et al. 2015), in principle I allow for three types of recognizing of emotions: two types of direct perception are distinguished in terms of top-down processes of shaping perception involving background images or beliefs; and one is characterized by theory-based inferences. Thus, I distinguish "(1) (a basic form of) perceiving an emotion in the (near) absence of any top-down processes, and (2) perceiving an emotion in a way that significantly involves some top-down processes (a strongly concept-modified form of perception). Both types of perceiving emotions can be distinguished from (3) inference-based evaluation of an emotion pattern. The latter presupposes a stable evaluation of an emotion as being F, which then may be modified or reevaluated by reflecting on the information" (Newen et al. 2015, p. 197). To sum up: Direct perception can be based on a metaphysical framework that regards emotions as integrated patterns of characteristic features and this allows me to combine it with presupposing mental representations of emotions (as memorized rich patterns), on the one hand, as well as with a non-inferential recognition of some emotional episodes on the other. The pattern theory of emotion is furthermore able to account for internalistic features of emotions like the feeling of fear, but also for individualistic yet expressive features like behavior and expression in face, gesture, and body posture. This metaphysics of emotions is coherent and is compatible with several epistemic strategies for recognizing them, e.g. direct perception as well as theory-based inferential understanding.

Let me make a further clarificatory remark about my reply to the coherence worry: I illustrated my metaphysics taking emotional episodes as a core example. This does not imply that I analyze *all* mental phenomena in this way. Although I think that some mental phenomena can also be individuated as integrated patterns of characteristic features like self-awareness/self-consciousness (see Gallagher 2013) or object perception, I remain neutral on the question of how far this analysis can be generalized and about the possibility that some mental phenomena need a different metaphysics as basis. For this reply it is sufficient to have shown what a concrete paradigmatic example of a coherent metaphysics for emotional episodes looks like, in order to prevent the danger of running into an incoherent metaphysics as a unavoidable consequence of the multiplicity view concerning epistemic strategies of understanding others.[2]

4 Quadt's proposal FOR an alternative metaphysical framework

Although I think I do not need an alternative metaphysics, since I have a coherent one already, I would like to briefly comment on Quadt's account. She starts with a remark on embodiment. I do not really see any serious disagreement with my views here. For it is fine by me that phenomenal properties and mental representations in general are realized within the body —and sometimes not only in the brain but within our whole body (see the discussion of emotions). Furthermore, I said that in this reply I leave open whether we need an extended realization basis for some mental representations. Quadt's alternative proposal, with which she aims to deliver a new framework for a multiplicity view, introduces different levels of embodiment. One way to read her distinction is that it offers a characterization of different types of representation that unfold during ontogeny. This basic idea is entirely consistent with my work. In other papers I discuss in detail the development of different types of representation in ontogeny (Newen & Vogeley 2003; Newen & Fiebich 2009; de Bruin & Newen 2012). There are of course differences in how one might form types of representation but discussion of these goes beyond the scope of this reply.

Let me now elaborate on an important point of disagreement. Quadt's proposal is based, among other things, on the distinction between transparent and opaque ways of being involved in a mental state. She takes this distinction from Metzinger (2003, 2004). We can illustrate this dis-

2 Let me highlight that the multiplicity view of understanding others is only one part of my person model theory and this epistemic aspect is in addition defended and further developed by my former PhD-student Anika Fiebich in the following paper which just appeared: Fiebich & Coltheart 2015.

tinction using the example of the mental event of perceiving an apple. This event is transparent if I am only consciously aware of the apple, while it is opaque if I am (also) aware of my mental state of seeing the apple: "[w]hat distinguishes transparent from opaque states is the degree to which one's own social cognitive processing, which is directed at the other person, is explicitly represented as a process" (Quadt 2015, p. 12). The relevant move is Quadt's claim that the epistemic access of direct perception in social cognition can be explained by transparency, while the epistemic access of simulation and theory-based inference can be explained by opacity.

Here I think she is on the wrong track. This distinction between transparency and opacity in the case of a mental state of attributing a belief leads to the idea that I am not only aware of the other person having a belief with content p but that I am also focussing on being consciously aware of the process of my attributing a belief to the other. The latter can of course happen in case of reflective processes of attributing beliefs; but normally we are in a mode of just using our ability to attribute beliefs automatically, focusing on the other's belief and its content (not on our own process of attributing it). We normally deal with our mental state of attributing beliefs in a transparent way, contrary to the analysis offered by Quadt. Furthermore, direct perception can also be used opaquely in rare cases of being reflectively aware of guiding images: if I am an experienced chess player, I can perceive the chess board in a way that is best described by cognitive penetration, and in some cases I may be aware of the mental image which guides my perception, i.e. I see a position and know how to act because I consciously memorize the fact that I see exactly the same position I saw in a previously played game. Thus, the distinction between transparency and opacity is not helpful for characterizing the different strategies of epistemic access to another's mental states.

5 Self-models and person models: how are they related?

Finally let me point out an important question raised by Quadt, namely how are person models and self-models related to each other? A self-model is a special type of person model, the person model that someone develops of herself. This is also done at the two levels of an implicit self-schema and an explicit self-image. I intend to elaborate on the interaction between self-models and person model of others in future articles, but I completely agree with Quadt when she says that there is bi-directional informational exchange regarding both types of models in humans (which is also indicated in my paper in figure 2, p. 21): "I thus conclude that it should not only be considered how the development of a self-model influences social cognition, but also which role social processes play in forming such a self-model" (Quadt 2015, p. 10). The PMT has potential as a framework for a theory of human self-consciousness.

References

Barlassina, L. & Newen, A. (2013). The role of bodily perception in emotion: In defense of an impure somatic theory. *Philosophy and Phenomenological Research*, 1-42.

Clark, A. (2015). Embodied prediction. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

de Bruin, L. & Newen, A. (2012). An association account of false belief understanding. *Cognition*, *123* (2), 240-259. 10.1016/j.cognition.2011.12.016

Ekman, P., Friesen, W. V. & Ellsworth, P. (1972). *Emotion in the Human Face.* Oxford: Pergamon Press.

Fiebich, A. & Coltheart, M. (2015). Various ways to understand other minds. *Mind and Language.*

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

——— (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17* (2), 535-543. 10.1016/j.concog.2008.03.003

——— (2013). A pattern theory of self. *Frontiers in Human Neuroscience*, *7* (443), 1-7.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading.* New York, NY: Oxford University Press.

Gopnik, A. & Meltzoff, A. N. (1997). *Words, Thoughts, and Theories.* Cambridge, MA: MIT Press.

Hohwy, J. (2013). *The Predictive Mind.* Oxford: Oxford UP.

——— (2015). The neural organ explains the mind. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Macpherson, F. (2012). Cognitive penetration of colour experience. Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, *84* (1), 24-62. 10.1111/j.1933-1592.2010.00481.x

Metzinger, T. (2003). Phenomenal transparency and cognitive self-reference. *Phenomenology and the Cognitive Sciences*, *2*, 353-393.

——— (2004). *Being no one: The self-model theory of subjectivity.* Cambridge, MA: MIT Press.

Newen, A., Welpinghus, A. & Juckel, G. (2015). Emotion recognition as pattern recognition: the relevance of perception. *Mind & Language*, *30* (2), 187-208.

Newen, A. & Fiebich, A. (2009). A developmental theory of self-models. In W. Mack & G. Reuter (Eds.) *Social Roots of Self-Consciousness. Psychological and Philosophical Contributions* (pp. 161-186). Berlin, GER: Akademie 2009.

Newen, A. & Vogeley, K. (2003). Self-representation: searching for a neural signature of self-consciousness. *Consciousness & Cognition*, *12* (4), 529-543. 10.1016/S1053-8100(03)00080-1

Noë, A. (2005). *Action in Perception.* Cambridge, MA: MIT Press.

O'Regan, J. K. & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24* (5), 939-1031.

Quadt, L. (2015). Multiplicity Needs Coherence – Towards a Unifying Framework for Social Understanding. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Siegel, S. (2012). Cognitive Penetrability and Perceptual Justification. *Nous*, *46* (2), 201-222.

Vetter, P. & Newen, A. (2014). Varities of cognitive penetration in visual perception. *Consciousness and Cognition*, *27*, 62-75.

Welpinghus, A. & Newen, A. (2012). Emotion und Kultur. Wie individuieren wir Emotionen und welche Rolle spielen kulturelle Faktoren dabei? *Zeitschrift für philosophische Forschung*, *66* (3), 367-392.

Zinck, A. & Newen, A. (2008). Classifying emotion: a developmental account. *Synthese*, *161*, 1-25.