

UNDERSTANDING OTHER MINDS:  
A CRITICISM OF GOLDMAN'S SIMULATION THEORY AND  
AN OUTLINE OF THE PERSON MODEL THEORY

Albert NEWEN & Tobias SCHLICHT  
Ruhr-Universität Bochum

*Summary*

What exactly do we do when we try to make sense of other people e.g. by ascribing mental states like beliefs and desires to them? After a short criticism of Theory-Theory, Interaction Theory and the Narrative Theory of understanding others as well as an extended criticism of the Simulation Theory in Goldman's recent version (2006), we suggest an alternative approach: the *Person Model Theory*. Person models are the basis for our ability to register and evaluate persons having mental as well as physical properties. We argue that there are two kinds of person models, *nonconceptual person schemata* and *conceptual person images*, and both types of models can be developed for individuals as well as for groups.

Consider Ralph. Ralph is strolling along the beach where he sees a man wearing a brown hat, black sunglasses and a trench coat. He has seen this man several times before in town and his strange and secretive behaviour has made Ralph suspicious. Since the man, let's call him Ortcutt, always tries to cover his face and turns around all the time to see if he is being followed etc., Ralph has come to believe that Ortcutt might be a spy. Since Ralph finds this exciting, he follows him. Now, Ortcutt is in fact a spy and when he turns around and notices Ralph, he starts walking faster, takes his cell phone out of his pocket and makes all kinds of wild gestures while talking to someone. Ralph, in turn, comes to believe that the man in the brown hat believes that Ralph has recognized him as a spy and that his cover has been blown. Only now does it occur to Ralph that it might not have been such a good idea to show so much interest in the man and he runs away. How does Ralph acquire this belief about what Ortcutt might be thinking? This question is an instance of the more general question of how we understand others, how we come to know what they believe and

desire or intend to do, what they feel and perceive. Typically, when we think about what others are (or might be) thinking, we *represent* them as having mental states (processes or events<sup>1</sup>) like beliefs, desires, emotions and so on. This mental capacity of ours is sometimes called mentalizing or mindreading and it has been among the most-discussed topics in recent philosophy of mind and cognitive science. This research area has been transformed profoundly by recent developments in the cognitive neurosciences and developmental psychology. In the last decade, there has been an intensive investigation into the neural mechanisms underlying the capacities associated with mindreading and we have also learned a lot about some of the relevant capacities displayed by young children at various ages. Thus, research in this field has become essentially interdisciplinary. But despite the scientific progress in the empirical disciplines, there is still no consensus about how we should best understand and conceptualize these capacities subsumed under the name of mindreading. What exactly do we do when we try to make sense of other people by ascribing mental states like beliefs and desires to them? How should we best characterize the mechanisms of this capacity being executed *in us* when we do this?

1. *Theory-Theory, Simulation Theory, Interaction Theory, and Narrative Theory*

Four systematic positions can be distinguished under which most theories that are currently on the table can be subsumed (while some accounts are hybrids of these approaches). According to the so-called ‘Theory-Theory’ (TT), when we ascribe mental states like beliefs and desires to a person, we employ a folk-psychological theory similar to a scientific theory (e.g. Gopnik and Wellman 1994, Gopnik and Meltzoff 1997). Without such a theoretical embedding we cannot make sense of other people’s behaviour. This idea stems largely from experiments showing that children gradually learn about people and start to explicitly represent other people’s propositional attitudes at around the age of four years, when they are capable of ascribing *false* beliefs to others (Wimmer and Perner 1983). Until that age, children have acquired mental state concepts by observing others and thus have formed such a (rudimentary) theory, which may, on the

---

1. In general, we will use the notions ‘mental state’, ‘mental process’, ‘mental event’ interchangeably and do not want to make any ontological commitments regarding these notions.

basis of new observations, be revised during their cognitive development, just like a scientific theory may be revised given new observations. On this view, Ralph is like a scientist, trying to make sense of his observations by positing mental states as theoretical entities, and ascribing them to Ortcutt's mind—just like a scientist may posit theoretical entities like quarks and strings to explain certain observations. A competing version of Theory-Theory is based on a modular approach to the mind; it distinguishes various innate modules and claims that one such specific innate mechanism in our brain is designed particularly to understand other minds (a modular version of the Theory-Theory is defended by Baron-Cohen et al. 1985, Baron-Cohen 1995, Leslie 1987). On this view, Ralph employs this innate mechanism in order to understand Ortcutt's behaviour. What these approaches have in common is the contention that we employ a rather detached theoretical stance towards people, analogous to scientists who employ a theoretical stance towards their subject matter.

An alternative approach is the 'Simulation-Theory' (ST) put forward in different versions by Gordon (1986), Heal (1986), Goldman (1989, 2006) and others. The central tenet of this theory is that we use our own experience as an internal model, i.e. we *simulate* in our own minds what the other person might be thinking. Thus, we explore the mental states of others by putting ourselves in the position of the other in a current situation. We create pretend mental states in ourselves, which we then ascribe to or project onto the other. On this view, Ralph does not employ a theoretical stance towards Ortcutt but uses his own mind as a model and puts himself in Ortcutt's 'mental shoes' in order to find out what he thinks.

A third and more recent approach is called 'Interaction-Theory' (IT), defended by Gallagher (2001, 2005) and others (see also Gallagher and Zahavi 2008, Hobson 2002, Ratcliffe 2007, Reddy 2008). It stands in strong opposition to the first two approaches in rejecting a crucial assumption shared by those two rival views, namely, that there is even a problem of gaining *access* to other people's minds in the sense that they have mental states which are 'hidden' behind their behavior, while the latter is always everything that we can observe. It is rejected that we only have access to meaningless behavioural patterns and only subsequently hypothesize that this behaviour is guided by mental states. Proponents of this view emphasize that, on the contrary, we are typically engaged in second-person conversational situations with others whom we share a world with. In such social interaction, we mostly play an active part ourselves instead of taking

a detached theoretical stance towards the other. Such pragmatic interaction is, according to Gallagher, to a large extent characterized not only by what people are saying, but also by their embodied practices, including bodily movements, facial expressions, gestures, and so on. The central claim of this view is that we can, at least in most cases, *directly perceive* what other people are up to; neither theoretical inference nor simulation are thus the most pervasive ways of understanding others, they are seldom necessary.<sup>2</sup> Thus, according to this view, Ralph can somehow directly perceive what Ortcutt is up to. His beliefs and desires can be ‘read off’ his behaviour on the basis of Ortcutt’s embodied communicative practices such as displaying nervous movements, turning around many times, covering his mouth with his hand while talking on the phone and so on.

Another recent development is Hutto’s (2008) so-called ‘Narrative Practice Hypothesis’, which states that from the beginning of childhood we are exposed to and engage in various narrative practices; in direct encounters but also in various other situations we are exposed to stories about people acting for reasons. Such stories form the basis of our acquisition of the forms and norms of folk psychology. Thus, Ralph may understand Ortcutt’s behaviour on the basis of his (stereotypical) knowledge about spies, which he may have acquired via the relevant stories, e.g. from reading novels or watching movies.

We do not claim that this is an exhaustive list of positions that one may develop on this issue, but they are the ones that have been discussed extensively in the recent literature and distinguishing between them suffices for the purposes of this paper. The bulk of this paper is devoted to a critical discussion of the Simulation-Theory of mindreading. More specifically, we will focus on Alvin Goldman’s recent elaborate defense of this theory (Goldman 2006, forthcoming). To characterize the main criticism right at the beginning it is helpful to distinguish two demands: 1. We can ask which mental states someone else might have and how we come to know about this. 2. We can try to estimate the decision someone is going to make presupposing knowledge about the other person’s initial mental states (especially the relevant beliefs and desires). We argue that Goldman’s theory of high-level mindreading focuses only on the second question and thereby only deals with a very special case of understanding other minds which cannot be generalized. This case misses the main task

---

2. For a critical discussion of the notion of ‘direct perception’ in this context see Van Riel 2008 and Gallagher 2008a, b.

of a theory of understanding other minds (without additions it simply involves the mistake of a *petitio*) since the simulation of decision-making already presupposes an initial understanding of the other's beliefs and desires. In his recent approach, Goldman introduces a theory of low-level mindreading which deals with the relevant question 1, but as we argue below, (i) it cannot account for almost all propositional attitudes and (ii) it is not clear why it should be evaluated as being a case of mental simulation. Therefore, Goldman's Simulation Theory suffers from severe gaps given that it wants to offer a complete theory of understanding other minds. We grant that it is an important progress that he introduces the distinction between 'low-level' and 'high-level' mindreading as two radically different ways of understanding others. Our positive account will benefit from it. After offering a detailed characterization of Goldman's theory in section 2, we put forward several objections to his approach (sections 3 and 4), where section 3 is devoted to low-level mindreading and section 4 concerns high-level mindreading. We argue that the two accounts suggested by Goldman are so essentially different in kind and in complexity, that it is unmotivated to subsume both of them under the same umbrella of a generalized Simulation-Theory. It is explanatorily more fruitful to accept a multi-level theory of understanding other minds, based on the insight that we have very different strategies and mechanisms at our disposal for understanding others. Whether and when we employ these various strategies depends not only on our prior relation to the person whose 'mind' we wish to understand, but also on their behavioural patterns which we observe and on the context of the situation in which the observed person displays these patterns. Thus, we need a new alternative account in order to capture all cases of understanding others. In section 5, which is the constructive part of the paper, we suggest that we essentially rely on 'person models' to understand other minds. We introduce and explain this notion and distinguish two different kinds of person models: person schemata and person images. Person schemata are sufficient to establish a non-conceptual understanding while person images are constitutive for a conceptual understanding. Person models in general are used for self-understanding as well as for understanding other minds.

## 2. Goldman's Simulation-Theory

### 2.1 The general structure of Goldman's theory

When evaluating the alternative accounts of mindreading mentioned above, one needs to keep in mind that they do only exclude each other if each of them is interpreted as making the strong and universal claim that only one of them is *the single* (or at least *the most pervasive*) strategy we use to understand others (Cf. e.g. Baron-Cohen 1995, 3f, Goldman 2002, 7f). Indeed, it seems that if proponents of these various approaches would not make this strong claim, then there might not even have been such a lively debate in the past twenty years or so. Once one allows for different kinds or strategies of mindreading, both simpler and more complex ones, then also hybrid accounts combining elements of some of them are possible. Goldman defends such a hybrid theory, “a blend of ST and TT, with emphasis on simulation” (2006, 23). One of the reasons why he no longer subscribes to a pure Simulation-Theory is the phenomenon of self-ascribing current mental states for which the simulation routine just does not make sense.<sup>3</sup> In order to highlight the essential structure of the Simulation-Theory, it helps to contrast it with the structure of the *Theory-Theory*. Here, it is important to note that Goldman discusses the differences between these two main rival views only in the special context of predicting a decision, i.e. of someone's prediction of what another person shall decide on the basis of given beliefs and desires. As already mentioned, Goldman owes us a story how we come to know the initial propositional attitudes while the Theory-Theory explicitly accounts for them:

It is an essential ingredient of Theory-Theory that the attributor employs a background belief in a folk-psychological law, e.g. a law about means-end reasoning. For example, Ralph may run away since he believes both that Ortcutt has the initial belief that he has been exposed by someone, that he desires to get rid of this person and that (generally) ‘in situations where their cover is blown, spies usually decide to consult a colleague or their boss to ask them whether they should kill the guy who blew their cover’.

---

3. Another reason is that he accepts that Simulation-Theory cannot account for our understanding of other minds in the numerous cases of people suffering from mental diseases, which involve radically different experiences (e.g. thought insertions in schizophrenia, the experiences which are connected to Cotard Syndrome, and so on). It will be argued below that the additions necessary to account for such phenomena radically change the Simulation-Theory such that it is no longer adequate to characterize it in the intended way.

Ralph's beliefs about Ortcutt's initial mental states result from wondering about how to make sense of the target's behaviour. The target's presumed beliefs are treated like hypothetical theoretical entities and are in turn fed into a reasoning-mechanism, which then yields further beliefs (or rather, an inference) as output. The result is first the attributor's belief that the observed person is a spy and that he noticed that his cover has been blown. Then the attributor uses his reasoning mechanism to infer that the target person, in this case Ortcutt, decides to kill him.

According to Goldman, ST just presupposes the same initial mental states but it tells a different story about *how they are used* by the attributor: The attributor uses the "information that T desires g ... to create a pretend desire" (Goldman 2006, 28). Similarly, the attributor creates pretend beliefs, which are supposed to match the target's initial beliefs. These pretend mental states are then fed into the attributor's own decision-making mechanism resulting in a pretend decision, which, crucially, does not result in an action. Instead of being carried out or acted upon, this (pretend) decision leads to a *genuine* (not pretend) belief about what the target will decide to do in this situation. Thus, on this account, Ralph asks himself what he would do if he faced Ortcutt's situation and thus creates the *pretend belief* that his cover has been blown and the *pretend desire* to get rid of the man who exposed him, only to reach the *pretend decision* to kill this man. Then, instead of acting upon this decision, he *projects* it onto Ortcutt.

This schema characterizing ST has the following important features: First, the pretense involved in the creation of pretend propositional attitudes is a special kind of imagination. In contrast to imagining *that* something is the case, e.g. *that* someone is elated or that one sees a car, one imagines *feeling* elated or *seeing* a car. That is, one creates a state that is phenomenologically more similar to the real feeling or perception since one *enacts* the relevant state. Therefore, Goldman calls the relevant kind of pretense 'enactment imagination'. It involves a deliberate creation of a mental state with a special phenomenal character (Goldman 2006, 149). This state is then *projected* onto the other subject. A further feature of the mindreading process is the process of "quarantining". In order for the simulation routine to work it is crucial that the attributor's own mental states do not interfere with the pretend states. Thus, in the example, Ralph needs to "quarantine", i.e. isolate or 'repress' his own idiosyncratic beliefs and desires (Goldman 2006, 29). Failing to do so may result in an egocentric bias that contaminates the evaluation of Ortcutt's mental states. Thus,

according to Goldman, third-person attribution of a decision (high-level mindreading) consists of

- (i) creating pretend propositional attitudes (in a special way through enactment imagination)
- (ii) using a (the same) decision making mechanism (as in the first-person case)
- (iii) projecting the product of this decision-making process onto another person (attributing the decision), while quarantining those mental phenomena that are only specific for me and not for the other person.

Goldman tries to generalize this model to account even for basic forms of understanding other minds while introducing some modifications. In general, Simulation-Theory can be distinguished *negatively* from Theory-Theory by the rejection of the belief in a psychological law (or generalization) posited by TT, but it can also be *positively* characterized by positing this two stage-process of mindreading, namely the simulation stage and the projection stage (Goldman 2006, 40). The simulation stage demands a process P in the attributor that duplicates, replicates or resembles the relevant process P\* realized in the person observed and it should always result in a first-person-attribution of a mental state. The second stage is then the projection of this type of mental state onto the other subject.

## 2.2 *Low-level and high-level mindreading*

Let us critically examine these core features while adopting Goldman's useful distinction between low-level and high-level mindreading. Mindreading in general comprises all cases of evaluating the mental state(s) of another person, including the language-based attribution of a mental state to a person. Now, in the last section, the general pattern of mindreading postulated by ST has been introduced with a focus on propositional attitudes like beliefs and desires, and on the prediction of a decision made by someone else. According to Goldman's distinction, this is a typical case of high-level mindreading, to be contrasted with low-level mindreading. The latter is defined as a process which is "comparatively simple, primitive, automatic, and largely below the level of consciousness" (Goldman 2006, 113). It typically targets relatively basic mental states like emotions, feelings, sensations like pain, and basic intentions and it is usually grounded

in basic perceptual information. A paradigm case of low-level mindreading is thus face-based recognition of emotion. According to Goldman, such low-level mindreading is based on a mirroring process “that is cognitively fairly primitive” (ibid.). Thus, low-level mindreading may be caused or generated by the activation of ‘mirror neurons’. These neurons, which have been discovered about ten years ago in macaque monkeys, are activated both when the monkey *executes* a goal-directed hand action (reaching for and grasping a peanut, say) *and* when the monkey *observes* another individual (be it a monkey or a human being) executing a similar action (Rizzolatti et al. 1996, Gallese et al. 1996, Rizzolatti and Craighero 2004). According to Goldman, in order for a genuine mirroring process to take place, it is not enough that mirror neurons be activated endogenously. This may be the result of mere accidental synchronisation. Instead, they have to be activated *in an observation mode*, which excludes imagination-based mirroring, like motor imagery, from counting as a case of mindreading (cf. Goldman forthcoming). Therefore, although mirroring alone does not constitute mindreading, low-level mindreading may be based upon it or caused by it. To mention only one empirical example, it has been shown that activating a specific neural circuit underlying the *experience* of disgust is also causally efficacious in the normal *recognition* of this emotion in others, while failing to activate it (because of a brain lesion, for example) prevents both the capacity to experience it and the capacity to recognize it in and attribute it to others (Wicker et al. 2003).

A mirroring event needs to be supplemented by a classification of the target’s mental state(s) and a projection (or imputation) of that classified state onto the target. Although a case of mindreading demands both a simulation and a projection stage, the simulation stage need not involve multiple steps, but may be constituted by a “single matching (or semi-matching) state or event” (Goldman 2006, 132).

But not all mindreading is caused by or based upon mirroring, as Goldman emphasizes. This is so partly because “some forms of mindreading are susceptible to a form of error to which mirror-based mindreading isn’t susceptible” (Goldman forthcoming). Such errors are typically egocentric “failures of perspective-taking” or inhibition of self-perspective which simply cannot happen in mirroring. Secondly, the definition of a ‘mirroring process’ explicitly excludes imagination-driven events, while mindreading can sometimes be initiated by the imagination (e.g. when one learns about the other person’s situation from an informant). High-level mindreading then is defined as follows:

'High-level' mindreading 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, 147)

Thus, high-level mindreading is paradigmatically illustrated by the evaluation of a decision someone is going to make, as the example above illustrated.

First of all, we wish to emphasize that we applaud Goldman's intention to introduce a distinction between two such radically different kinds of mindreading instead of trying to account for all cases of mindreading by mentioning one single mechanism or strategy. The problem is that his way of drawing this central distinction is rather sketchy and ultimately does not withstand close scrutiny. For example, it is not clear whether the relevant criterion is the type of mental state to be attributed (a sensation or a propositional attitude) or whether it is the fact whether the mindreading process is conscious or not. Moreover, Goldman merely demands that in the case of high-level mindreading 'one or more' of the relevant features are present. He apparently does not intend to introduce necessary or sufficient conditions, and it is disputable whether he points out adequate conditions. A further problem is that although he mentions these crucial differences between low-level and high-level mindreading, Goldman still claims that both are essentially cases of simulation and that they can thus both be accounted for by his two-stage framework of 'simulation plus projection'. In the following two sections, we take issue with both the way Goldman draws the distinction in the first place and with his interpretation of these two strategies of mindreading as cases of simulation by looking more closely at both strategies, starting with low-level mindreading.

### *3. Problems of the Simulation-Theory of low-level mindreading*

As has been explained above, low-level mindreading is supposed to proceed in two steps, a first step of registering a type of mental state, e.g. an emotional or painful experience and a second step of projecting the emotional or sensational state in question onto another subject. Registering the emotional state is supposed to be constituted by a mirroring process on the neuronal level. Mirroring another person's emotional state amounts to the activation of the same neurons in the observer's brain, which would

be activated in case the observer felt the emotion or pain herself. This mirroring process is not something being subject to conscious control of the observer. Rather, it happens automatically and remains unconscious. In order for this to be a case of mindreading, a second step needs to follow. The observer needs to attribute the emotion or sensation in question to the other. This cannot be done unconsciously; it is rather a conscious and deliberate action. According to Goldman, the attribution of the mental state to the other person involves *projection*. He suggests that in every case of understanding others we first detect the mental state as a state of ourselves, secondly attribute it to ourselves, and then thirdly project it onto the other person. More explicitly, we find the following steps in Goldman's account of low-level mindreading (see Goldman 2006, 128):

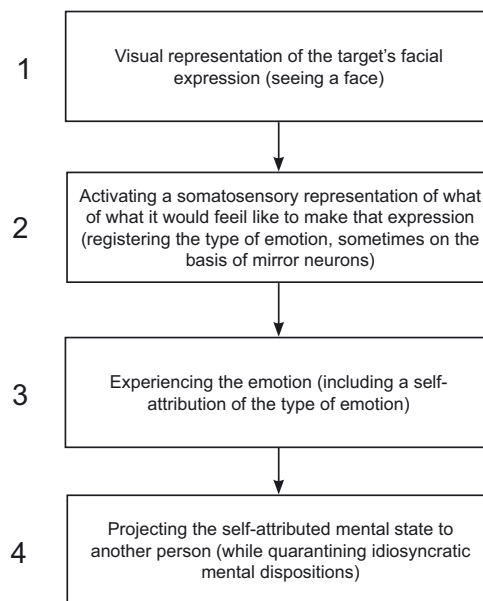


Figure 1

Let us now critically examine the three steps 2,3,4 of realizing simulation and projection in cases of low-level-mind-reading: We all agree what the minimal basis of low-level mindreading is: in the case of underlying mir-

ror neuron activation we register a type of mental state, e.g. tooth pain, independently from representing the person having it: “A simulation process can consist, minimally, of a single matching (or semimatching) state or event.” (Goldman 2006, 132) Here, our critical remark is why we should still characterize this kind of registering as a case of simulation. It is radically different from simulation in the case of high-level mindreading since the crucial element of enactment imagination is missing and the generalized condition of ST, namely that the representation “duplicates, replicates or resembles the mental process P in the other person in some significant respects” remains radically underspecified. Since the mirror neuron processes are unconscious, the “significant respects” cannot involve any conscious features of mental phenomena. So then the candidates of simulation are processes sharing the functional role of the unconscious automatic processes underlying the mental phenomena of another person. Simulation is then reduced to a resembling representation which does not involve any similar conscious experience or any state of pretending. It is not useful to summarize both processes in the cases of high-level and low-level mindreading under the same label of “simulation”. Gallagher suggests that we should best interpret mirror neuron activation in terms of direct perception (Gallagher 2007). After all, being in ‘observation mode’ is part of Goldman’s (Goldman, forthcoming) definition of a mirroring process. One may take that literally and argue that in many cases we can simply observe, i.e. ‘perceive’ other people’s mental states; we can just ‘see’ them in their embodied practices (gestures, facial expressions, etc.). For example, we can often see that someone is disgusted or in pain simply by looking at their facial expressions. Why would we need to posit a simulation process? Gallagher’s (2007) alternative perception-based account is more parsimonious and persuasive here. Furthermore, in cases where we have not yet experienced the relevant mental state ourselves we still start to create an attribution of a mental state. It seems that in such cases, Simulation-Theory has nothing to say. In such cases, other strategies may be needed.<sup>4</sup>

---

4. Goldman may try to treat such cases as exceptions, which he can account for since he defends a hybrid of ST and TT. Here, a strategy of theoretical explanations seems to be relevant. The problematic presupposition in this reply is that it involves—without sufficient reason—the claim that those cases are exceptions. We will argue in our constructive part (see section 5) that we as adults regularly have to attribute mental states that we do not experience ourselves. Otherwise we cannot understand the majority of the people in sufficient detail. Here we may have to switch to high-level mindreading even in the case of emotions and sensations.

Goldman's step 2 also underestimates the fact that the mirror neurons only represent a type of mental state without being sufficient for a self-other distinction. The fact that mirror neurons fire irrespectively of whether the monkey executes the action or merely observes the other executing it suggests that this firing does merely encode an action plan but is otherwise completely neutral with respect to *who* is performing the action. Far from *solving* the problem of understanding others, the mirror neuron discovery seems to *give rise* to the question what *further* mechanism enables us to distinguish our own actions (or mental states) from those of others. It points to the need of some further system, sometimes called a 'Who-system' (Georgieff and Jeannerod 1998, de Vignemont and Fournier 2004) for registering a mental state as a state of ourselves (and not of someone else): The self-other representation is installed by a process at least partly independent from mirror neurons. This makes it plausible that the information about the type of the mental state is combined either with the self- or with the other-representation, but not with both. It may be part of the format in which an action plan is encoded that it is either first-personal (proprioceptive etc.) or third-personal (outer perception) and this difference in format might be realized in a different neural mechanism that interacts with the mirror system. Let us illustrate this with an example: If I see an angry face, my mirror neurons may be activated and represent the anger, but now the information that the activation is based on the visual input of a face automatically leads to an other-representation of the mental event. If such an other-representation of anger is established we only need to express the content linguistically in order to attribute it adequately to the other person. On this account, no intermediate linguistic self-attribution is needed. Even if we would grant that in all cases of observing mental states some experience is produced by mirror neurons inside of me, such an experience need not lead to a self-attribution of the mental state. But this is what Simulation-Theory claims when it posits that *in general*, understanding others proceeds by modelling the other's mental states with one's own first-person experience.

While Goldman concedes that simulation is radically impoverished in the case of low-level mindreading, he suggests that projection is the same in low-level and high-level mindreading. The case of low-level mindreading is also supposed to involve the self-attribution of the mental state (step 3 in his model) which then leads to an attribution to another person: "If, in addition, the observer's classification of his own emotion is accurate, his attribution of that same emotion to the target will also

be accurate.” (Goldman 2006, 129) What is the status of the proposed self-attribution? It seems that Goldman faces a dilemma here: If he claims that the self-attribution is a *conscious* event, then this stands in contrast to our phenomenological observations (and to his general characterization of low-level mindreading as an unconscious process): We simply do not consciously self-attribute pain or disgust in the case of observing someone else’s pain or disgust (at least in most everyday cases)<sup>5</sup>. But if Goldman claims that the self-attribution in low-level mindreading is an *unconscious* event, then we simply lack sufficient empirical evidence. We can offer an alternative explanation, which is more parsimonious and does not involve a projection on the basis of a self-attribution. As explained above, low-level mindreading is particularly manifested in the recognition of basic mental states like emotions, sensations (e.g. pain) and simple intentions (to grasp something, say). There is strong evidence that recognizing basic emotions like anger, fear, sadness, etc. on the basis of the perception of facial expressions is a strongly modularized process: if both amygdalae are damaged it seems to be impossible to experience fear and to register fear in other people. The relevant brain areas are ‘mirroring’ areas, underlying both the experience of fear and the registration of fear in others (Damasio 1999, 66, Goldman 2006, 115f.). But what is important here is that in order to describe the process of recognizing fear in another person—in normal cases—we just need to presuppose a self-other distinction in addition to the registration of fear. And such representations come in various degrees of complexity: A non-conceptual self-other distinction is already available for a cognitive system like humans (and other animals with a minimal behavioral complexity) on a very basic level of bodily self-acquaintance (Bermúdez 1998, Newen and Voegeley 2003, Newen and Vosgerau 2007, Vosgerau and Newen 2007). The combination of registering fear with a non-conceptual ‘other-representation’ is sufficient to register ‘fear in the other person’. To arrive at an attribution of fear, this other-representation of fear is expressed in natural language. Therefore, our alternative view ideally involves three closely connected elements of understanding an emotion in someone else: 1. the non-conceptual registration of the type

---

5. An exception may be a case where I am very closely related to the person who is suffering, e.g. if my child has burned her finger on the hot cooking plate. Although I may experience pain consciously in such cases, the most relevant experience is not pain but concern (about what to do next). Moreover, we would have to distinguish between real pain (on the basis of a burned finger) and mere emphatic pain (which is not caused by burning one’s finger). Also, it seems that in the case of contagion (for example, laughing or yawning) we have a third case to distinguish.

of emotion, 2. the combination of a non-conceptual other-representation with the non-conceptual registration of a type of an emotion, and 3. the expression of this content in natural language. No projection from self to other is involved in this model.

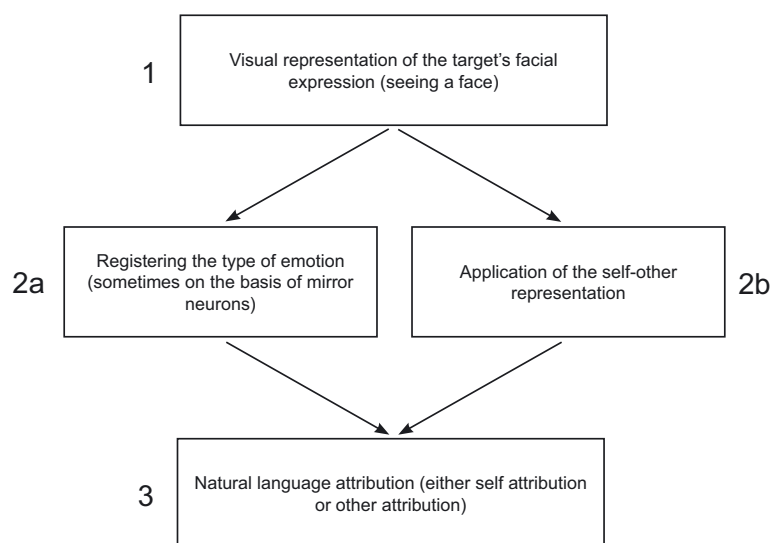


Figure 2

This is the standard scenario for understanding others on the basis of visual information. In most cases, visual information about a facial expression displaying an emotion triggers a representation of the type of emotional state in question and an other-representation, which then leads to an attribution of an instance of that type of emotion to the other person whose facial expression triggered the representation of the emotion. But note that visual information about a facial expression does not necessarily lead to an attribution of an emotion to *another* person. It may also lead to a self-attribution of that emotion, e.g. when one looks in a mirror and receives information about one's own facial expression. In that case, the visual information triggers the application of a self-representation instead of an other-representation, based on prior knowledge about mirrors and about ourselves.

If a simpler story is available, especially if it involves less demanding cognitive mechanisms, then the principle of parsimony can be employed against Goldman's Simulation-Theory of low-level mindreading, in favour of an alternative, perception-based approach.<sup>6</sup> Let's now turn to our criticism of Goldman's Simulation-Theory of high-level mindreading.

#### 4. *Problems of the Simulation-Theory of high-level mindreading*

According to Goldman, while low-level mindreading is fully automatic and usually applies to emotions, sensations and intentions, high-level mindreading crucially involves "enactment imagination" as a cognitively high-level activity which is at least potentially under our conscious control. First of all, Goldman conceptualizes pretending as an operation or process, not as a distinct mental attitude in addition to belief and desire (Goldman 2006, 47), because otherwise we could not make intelligible what a pretend belief or a pretend desire were supposed to be. More specifically, pretense is supposed to be a kind of imagination. Goldman distinguishes various kinds of imagining: One can imagine *that* something is the case, e.g. *that* someone is elated. One can also imagine *feeling* elated or *seeing* a car. That is, imagining something may not amount to a supposition but to "conjure up a state that feels, phenomenologically, rather like a trace or tincture of elation ... When I imagine feeling elated I do not merely suppose *that* I am elated; rather, I *enact*, or try to enact, elation itself" (ibid.). This is what Goldman calls "enactment imagination" and it is the crucial process underlying high-level mindreading, e.g. mental states that are projected towards another person are supposed to be the results of this process. We will argue in the following that Goldman's account of high-level mindreading, put forward as a hybrid account of ST and TT, is unpersuasive as long as it is supposed to be a variant of Simulation-Theory, simply because of the elements of Theory-Theory that it relies upon at various points.

The major problem with Goldman's Simulation-Theory of high-level mindreading is that it does not even get off the ground. In short, it cannot provide an explanation of how we come to attribute mental states to another person and this is the core of mindreading. Recall that mindreading may be understood in two ways: One may ask how we recognize mental

---

6. This also holds against Goldman's interpretation of mirror neuron activity in terms of simulation (Gallese and Goldman 1998).

states in others and one may ask how we estimate the decision someone else is going to make *on the basis of* (our knowledge about the person's) initial beliefs and desires. These are two different questions and since, arguably, the second presupposes an answer to the first, the first is at the core of mindreading. In presenting his account of high-level mindreading, Goldman *presupposes* that we already know the target's beliefs and desires. Thus, it presupposes what it is supposed to explain. Our criticism, in short, is that ST does not provide an answer to the first question but only to the second, and in doing so, *presupposes* an alternative account of what it means to recognize or understand the beliefs and desires of others. And in this regard, Theory-Theory is more attractive than Simulation-Theory, especially since theoretical assumptions enter Goldman's hybrid account anyway. Let us elaborate this objection in more detail.

Goldman's model of high-level mindreading concerns decision-making, i.e. it starts with the attributor's beliefs about the initial mental states the target supposedly has (Goldman 2006, 26-30). In elaborating his model, Goldman explicitly says that Theory-Theory and Simulation-Theory start with the same assumptions on the part of the attributor regarding the target's initial mental states. In both cases, the attributor thinks that the target has the belief that *p* and the desire for *g*. Goldman says that the two accounts only differ with respect to *how* the attributor *uses* these presumed mental states, or to what the attributor *does* with them. So, a crucial element of Simulation-Theory, which Goldman does not elaborate in any detail, is the initial "information that T desires *g*" (Goldman 2006, 28) which the attributor supposedly has at her disposal. In order for the attributor to create (in herself) the correct pretend mental states, she needs to know in advance which mental states the target is undergoing, i.e. what the target initially believes and desires. Obviously, only if the attributor knows that the target desires *g*, she can create a pretend desire for *g* instead of creating a pretend desire for *f*. But importantly, this first step already constitutes what needs to be explained, namely mindreading or mental state ascription (the first question posed above). Thus, the simulation routine can only get off the ground *given* some prior knowledge about the target's initial beliefs and desires. Arguably, Goldman needs to tell a story about how the attributor arrives at her beliefs about the initial beliefs and desires of the target. If he cannot tell such a story, then the simulation routine does not have any explanatory power by itself.

Therefore, the question arises of how this initial "information" acquisition should be spelled out. In order to make good his case for the claim that

a simulation routine is the essential ingredient in the mindreading process, Goldman would have to supplement his account with a story about how the attributor comes to hypothesize that the target has the initial desire for *g*, and this story would need to be formulated in terms of simulation. But as we submitted above, it seems that simulation cannot do the job since it always presupposes knowledge about mental phenomena that can then be pretended. The alternative theories do not seem to face this problem. A more theoretical explanation, for example, does not depend on this condition. According to Theory-Theory, the attributor comes to posit the specific beliefs and desires of the target on the basis of her observation of the target's behaviour (which she cannot make sense of merely on the basis of the pure perceptual information). These hypothesized initial mental states are evaluated against some folk-psychological "generalizations" (to avoid the term "law") in order to come up with a further hypothesis about the target's decision.<sup>7</sup>

At this point, it may be useful to briefly elaborate what the belief in a "theory" may amount to. Some proponents of Theory-Theory argued that it is akin to a scientific theory (Gopnik and Meltzoff 1997, Gopnik 1993). This gave rise to a number of objections and a heated debate about how cognitively demanding TT is; we agree that this is problematic since it is not even agreed upon what a scientific theory really is and because this may be too demanding when it comes to infants and their capacity to understand others. For our purposes, a theory may be understood as a systematically interconnected set of beliefs regarding a set of phenomena. The relevant class of phenomena in this context are mental states and the relevant set of beliefs can be characterized as a certain limited number of generalizations. The advantage of Theory-Theory over Simulation-Theory in the case of high-level mindreading is that it is designed to deal with the first of the two questions above, which Simulation-Theory does not even attempt to answer. The claim that Theory-Theory is more persuasive than Simulation-Theory can be further justified by emphasizing the extent to which Goldman relies on elements of Theory-Theory at other points.

---

7. If a proponent of Simulation-Theory wants to include knowledge of folk-psychological principles as well as the evaluation of mental states on the basis of observing behavior by using these principles into his account, then this is not a modest modification into a hybrid account since the application of these principles then does the essential work in the process of understanding other minds. Even if Goldman's model of decision-making were correct, it would just be a very special case of understanding others presupposing the important basic case (see below).

Goldman's account makes use of a further hidden premise before the projection stage. Before the attributor projects her own pretend states onto the target, she tacitly assumes that this particular target (but also other people in general) is "like her" in the relevant respects. That is, the attributor tacitly believes that people are equipped with the same decision-making mechanisms and arrive at pretty much the same decisions, given certain beliefs and desires. Otherwise, the attributor would have no justification (or weaker, motivation) whatsoever to assume that her own pretend belief, desire and decision—arrived at by enactment imagination—resembles or even matches the target's belief, desire and decision. So the attributor believes that the target is like her in relevant (cognitive) respects. What is the status of this belief? Because of its generality and universality, it seems reasonable to regard it as a belief in a generalization (if not a psychological law) about people, just as suggested above.

Moreover, it seems that this belief in the semblance of decision-making processes in the attributor and the target contains what others have called a "rationality assumption" (Dennett 1987). This rationality constraint enters the story because of the relation between the target's presumed initial belief and desire which are of course interrelated: It needs to be assumed that the attributor believes that, given the desire for *g* and the belief that action *m* will lead to *g*, one should *rationally* arrive at the decision to do *m*. Otherwise, she would neither be justified (motivated) in arriving at her own pretend decision herself, nor would she have reason to believe that the target should arrive at this decision; that is, the projection would be unmotivated. But seen in this light, it is unclear why Goldman so vehemently opposes what he calls "Rationality Theory" (Goldman 2006, 53-68). Apart from the fact that Goldman's account of Dennett's theory is at times unfair, it is clear that this 'Rationality Theory' does not claim that we *always* think that other people make rational decisions. As Dennett (1971) emphasizes, at a certain point we may have to give up the rationality assumption when we try to make sense of other people's behaviour and this behaviour does not fit our model of rational decision-making. It seems that not even Simulation-Theory can do without the attributor's assumption of a minimal rationality in her own case as well as on the side of the target. Otherwise it is impossible to predict what one may decide; at least one can never be sure if anything is possible and the target may be completely irrational.

Furthermore, Goldman introduces the process of quarantining, according to which pretence requires that one must isolate or 'repress' one's own

idiosyncratic beliefs and desires to account for the case that the attributor notices that he has a (radically or partially) different mind-set compared to the other person. But how can Simulation-Theory account for such a process? Here we marked a further feature, which can also be best understood as a theoretical component. To be able to notice that the other person has essentially different mental states (a different mind-set) than myself, I have to represent a minimal model of the other person and a minimal model of myself, and I have to compare both 'person models' to register relevant differences. This observation provides essential support for our positive account, which we shall call the '*person model theory*' (see section 5).

To sum up these points, it seems that at three points Goldman has to make use of theoretical assumptions that have nothing to do with simulation. First, at the beginning of high-level mindreading, he invokes "information" regarding the target's initial mental states that is supposed to be at the attributor's disposal; and it is plausible that this information is arrived at by some sort of inferential process as posited by TT. Secondly, after the alleged simulation routine has taken place, the projection of the pretend decision arrived at is based on the belief that this pretend decision *matches* the target's decision, which in turn presupposes the attributor's belief in the resemblance of self and other (in relevant cognitive respects). Finally, to account for the exclusion of idiosyncratic mental states of the attributor, Goldman has to presuppose "quarantining" which introduces a third theoretical component.

Goldman foresees the second criticism, formulating it as the objection that ST ultimately collapses into TT. Against this, he presents various replies. First, he says that it would not be a total collapse of ST into TT if a "theoretical box" was added to the story since the overall process would still be simulational in nature. But since we have identified not only one but three theoretical assumptions or inferences that have to be added to the story, one wonders why the intermediate stages of the mindreading process should be framed as simulations and why this should matter very much, given the theoretical assumptions.<sup>8</sup>

---

8. But Goldman also rejects the objection that unless the attributor believes that the target is relevantly like her, she would not be *justified* in attributing her own decision to the target. Goldman's reason for this rejection is that he wants to distinguish this question of justification from the question of how mindreading actually works. But one need not put the objection in terms of justification. Instead, one can ask what would *motivate* the attributor to arrive at her own (pretend) decision to do *m* on the basis of her pretend mental states rather than to arrive at the (pretend) decision do *n*. It seems that only the assumption of a rational link between the initial belief and desire can make this intelligible.

Goldman's second major reply is to reject the claim that the belief in resemblance-to-self amounts to a belief in a psychological law (Goldman 2006, 31). If it is not a belief in a psychological law then the proponent of TT arguably has no argument for his position. Again, this dispute boils down to the debate about what counts as a "theory", according to Theory-Theory. As we have hinted at above, it is not easy to settle this dispute, partly because not even philosophers of science have a clear and uncontested answer to this question. We offered a minimal proposal, namely, that a theory consists of a set of beliefs about a class of phenomena. If one grants that the number of beliefs required for something to be a theory can be very small, then the crucial beliefs that Goldman needs to posit in his own account justifies to call it a belief in a theory. But when the fate of Simulation-Theory is at stake, it is not so much the point whether we want to call the relevant beliefs a psychological law or a theory. The point is rather that these beliefs are further crucial ingredients of the whole story and that they are not explained in terms of simulation.

A final problem arises if we take a closer look at the underlying neural mechanisms in the case of high-level mindreading: High-level mindreading in general is supposed to be based on the so-called 'mindreading network', which includes the medial prefrontal cortex and the temporo-parietal junction (Fletcher et al. 1995, Gallagher et al. 2000, Frith and Frith 2003, Saxe and Kanwisher 2003). But we have to distinguish between first-person and third-person attribution. According to Simulation-Theory, we would expect that the neural correlate of third-person attribution includes all those brain areas that are activated in the case of first-person attribution of mental states since self-attribution constitutes an essential stage in the simulation-and-projection process. But recent empirical evidence does not support that expectation: In the study of Voegeley et al. (2001) it has been shown that first- and third-person attribution have different neural correlates and, most importantly, that the correlate for third-person attribution does not include the significant activations of first-person attribution (see also Voegeley and Newen 2002).

These points together raise the question what makes Simulation-Theory of high-level mindreading so attractive in the first place since other crucial elements of the overall account have nothing to do with simulation. Goldman's contention that the whole process is still essentially a simulation routine is just that, a contention. Moreover, Nichols and Stich (2003) have pointed out correctly that Goldman is in danger of using the term 'simulation' for too many processes, which are too different in kind to form a

theoretically interesting category. This objection is especially pressing when one puts low-level simulation and high-level simulation together in one category. It seems that an account of high-level mindreading that relies merely on theory will be simpler and more parsimonious than a hybrid account that combines simulation and theory. At this point it looks like Goldman's Simulation-Theory does not withstand closer scrutiny. While what Goldman calls high-level mindreading should best be explained in terms of the application of a theory, his low-level mindreading should best be explained in terms of perception (or 'registration').

##### 5. *The person model theory of understanding other minds. An outline*

Before we introduce our own positive alternative account, we shall now finally turn to our objections against the way Goldman draws his distinction between low-level and high-level mindreading. As mentioned above, we applaud his general intention to draw such a distinction since there is empirical evidence of a twofold system of understanding other minds (e.g. Olsson and Ochsner 2007). But Goldman leaves it largely unclear what his criteria are.<sup>9</sup> Even more importantly, we suggest that we have to apply the distinction twice over: We should first distinguish kinds of mental phenomena, and secondly strategies of understanding other minds. Let us begin by illustrating the distinction of different kinds of mental phenomena:

1. Concerning emotions, it is already commonplace to distinguish *basic emotions* (like joy, anger, fear, sadness), which do not involve any higher cognitive processes, from *cognitive emotions*, which essentially involve propositional attitudes (Zinck and Newen 2008). Ekman (Ekman et al. 1969) has shown that there are culturally universal facial expressions of basic emotions. Damasio (2003) also argues that we have to account for what he calls primary emotions. Disagreement only concerns the question which emotions exactly count as basic. We can define basic emotions by the underlying (relatively) modular processes that are independent from higher-order cognition (Zinck and Newen 2008) and we may generalize this to define basic mental phenomena in general (like colour vision which in the case of a local lesion leads to achromatopsia). Then pain, disgust and

---

9. De Vignemont (2009) also criticizes Goldman's distinction and points out some problems regarding the compatibility of the way he draws the distinction and the empirical evidence of two neural networks for mindreading.

fear are basic mental phenomena without higher cognition being involved since they are realized by modular brain activations like mirror neuron activation in cases of pain and disgust, amygdala activation in the case of fear etc. Shame, jealousy, love, envy, etc. are cognitive emotions since they essentially involve propositional (or relational) attitudes; e.g., in addition to a basic feeling, envy presupposes the belief that someone has a valuable object which I do not have, but really want to have yet cannot get, given my abilities and further conditions (Zinck and Newen 2008). We propose to apply this distinction regarding emotions to all mental states. That is, in general, we can distinguish *basic mental phenomena*, which are realized by modular brain processes, and *high-level mental phenomena*, which essentially involve propositional (or relational) attitudes.

2. The second distinction concerns the way of understanding someone's mental phenomena: Face recognition is a well-known modular process, which is essentially relying on activations in the 'fusiform face area' (Kanwisher 2001). This is a typical example of an unconscious modular process, which realizes a registration of the other person's face and since this representation is coupled with the detection of basic emotions, we have here a basic process of registering other minds independent from high-level cognition. Such a registration can then produce an adequate reaction which still does not involve any higher-order cognition: The position of a slightly bended head, for example, signals sympathy and is understood as such on an unconscious level by mere registration. The behavioural response of also signalling sympathy is caused by this registration of the other's mental condition (Frey 1999). One may dispute that this form of registering the other's mind already counts as a case of *understanding* the other's mind. Nevertheless, we suggest to classify it as mindreading while we wish to highlight that the relevant strategy involved is a *non-conceptual* form of understanding.

The alternative strategy is a *conceptual* way of understanding other minds, which essentially involves conceptual and propositional representations (the distinction between these kinds of representations is elaborated in Newen and Bartels 2007). This is what Goldman has in mind. That is, what he calls low-level and high-level mindreading are *both* cases of the conceptual way of understanding other minds because they always involve a linguistic attribution of mental states. We can account for Goldman's distinction by granting that there is a conceptual understanding of other minds which can be further distinguished into two forms according to the relevant mental phenomena: a conceptual understanding of basic

mental phenomena (low-level mindreading) and a conceptual understanding of high-level mental phenomena (high-level mindreading). But in addition, there are important *non-conceptual* forms of understanding others not captured by Goldman's distinction. With these distinctions at hand we can develop an outline of a new theory of understanding other minds.

### 5.1 *What are the central claims of this account?*

We suggest that we develop 'person models' from ourselves, from other individuals and from groups of persons. These person models are the basis for the registration and evaluation of persons having mental as well as physical properties. Since there are two ways of understanding other minds (non-conceptual and conceptual mindreading), we propose that there are two kinds of person models: Very early in life we develop non-conceptual *person schemata*: A person schema is a system of sensory-motor abilities and basic mental phenomena<sup>10</sup> realized by non-conceptual representations and associated with one human being (or a group of people), while the schema functions typically without awareness and is realized by (relatively) modular information processes. Step by step, we also develop *person images*: A person image is a system of conceptually represented and typically consciously registered mental and physical phenomena related to a human being (or a group of people). Person models are created for other people but also for myself.<sup>11</sup> In the case of modelling myself we can speak of a self-model which we develop on the non-conceptual level as a self schema and on the conceptual level as a self image.

A person schema is sufficient to allow newborn babies to distinguish persons from inanimate objects, manifested in neonate imitation, which is also sufficient for seven month old babies to separate persons from animals (Pauen 2000). We already mentioned the observation of non-conceptual

---

10. Mental phenomena include different ontological types: states, events, processes and dispositions. So, not only stable mental phenomena are included but also situational experiences (like tokens of perceptions, emotions, attitudes, etc.). In a more detailed explication of the theory it would be useful to distinguish situational person schemata (only stored in working memory) and dispositional person schemata (stored in a long-term memory). This has to be done in another paper.

11. 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 what he introduced as *body image* (Gallagher 2005, 24).

understanding of other minds by unconsciously registering someone's position of her head as signalling sympathy (Frey 1999).<sup>12</sup> Those registrations are part of a situational person schema which influences our interaction even though we are not consciously aware of it. On the basis of such non-conceptual person schemata, young children learn to develop conceptual *person images*. These are models of individual subjects or groups. In the case of individual subjects they may include names, descriptions, stories and whole biographies involving both mental and physical dispositions as well as manifestations. Person images are essentially developed not only by observations but also by telling, exchanging and creating stories (or 'narratives').<sup>13</sup> Person images presuppose the capacity to consciously 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 step when children acquire the so-called theory-of-mind ability (operationalized by the false-belief task, see Wimmer and Perner 1983).

Person schemata are closely related to basic perceptual processes. Therefore, we adopt Gallagher's view that we can sometimes just directly perceive mental phenomena, but take it to be true only for basic mental phenomena. Person images presuppose higher-order cognitive processes including conceptual and propositional representations, underlying a conscious evaluation of the observations. Here our background knowledge plays an important role to evaluate the mental phenomena. So on our view, the theory of direct perception is implausible for these complex phenomena.<sup>14</sup>

To sum up: The understanding of other minds is based on unconsciously established person schemata and consciously developed person images (if the latter are already established in the course of cognitive development) while both are normally closely interconnected.

---

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

13. This is the true aspect of the narrative approaches of understanding other minds mentioned above (e.g. Hutto 2008). But narratives are only one method to establish a person model. Representatives of the narrative approach underestimate other sources like perceptions, feelings, interactions etc. which often do not involve narratives.

14. This is acknowledged by Gallagher (2005) since he supplements his theory of direct perception with a narrative view akin to Hutto's (2008).

## 5.2 *What is the evidence for our view?*

As far as the non-conceptual understanding of other persons is concerned, an important ability is biological motion detection: We can see, just on the basis of point light detection of a movement, whether an observed person is a man or a woman, whether s/he moves happily or angrily (Bente et al 1996, Bente et al 2001). This is a very basic observational ability which allows us to register basic intentions of actions as well as basic emotions, without necessarily being conscious of it. Hobson and colleagues (Moore et al 1997, Hobson 2002) showed that exactly this ability to perceive a biological movement as displaying a certain emotion (like anger or happiness) is impaired in autistic children. They do not understand bodily movements as expressions of emotions. These examples illustrate the capacity of a non-conceptual understanding of other minds and they indicate that we (at least healthy subjects) can directly perceive these basic mental phenomena. To support the latter claim, we rely on the classical study by Heider and Simmel (1944) who showed that typical kinds of movements are immediately seen by us as intentional actions even if they are realized by geometrical figures. Furthermore, recent studies using that paradigm show that autistic patients characteristically lack this ability (Santos et al 2008). Direct perception of basic mental phenomena like basic intentions and emotions is a standard ability and lacking it has dramatic consequences for social interactions because then the person schemata essentially lack the standard information we normally receive.

Furthermore, there is empirical evidence that we not only develop person schemata but also person images. Here we simply rely on folk-psychological evidence, on the one hand, and the theory-of-mind ability on the other, allowing us to establish complex person images. We can develop person images of individuals but also of groups. Those person models of groups are also called 'stereotypes'. Stereotypes are an essential part of characterizing groups. One function of stereotypes is to provide an economical way of dealing with other persons (Macrae & Bodenhausen 2000). Besides minimizing cognitive effort they also play an important role in social identification: Situating oneself inside some groups but outside others seems to be a constitutive process of developing a social identity: It has been shown that even independently from competitive conditions we start to support in-group members (of a group we belong to) and disadvantage members of the out-group (Doise & Sinclair 1973, Oakes et al. 1994). The existence of stereotypes is also supported by recent studies which try to identify the

relevant neural correlates of stereotypes in social comparisons claiming that medial prefrontal cortex is essentially activated in these cases (Volz et al. under review). So there is not only evidence from folk psychology that we rely on stereotypes by classifying people but also some support from recent social neuroscience. Concerning person images of individuals, we all share the intuition that we develop a very rich and detailed image of people who we are very familiar with, our husband or wife or our kids, say. To treat such a specific image as an image of an individual can again be disrupted in pathological cases, e.g. either when a patient thinks that the person image of my brother can be instantiated in very different people (Fregoli's syndrome involves a too coarse-grained individuation of person models) or in the case of patients suffering from Capgras' syndrome. The latter have the delusional belief that one of their closest relatives, e.g. their wife, has been replaced by an impostor. They typically say things like 'this person looks exactly like my wife, she even speaks and behaves like my wife but she is not my wife' (Davies and Coltheart 2001); they insist on a too fine-grained individuation of person models such that no one can satisfy it due to a lack of a feeling of familiarity. Such pathological cases can be accommodated nicely within our general framework of person models. The general functional role of person models is to simplify the structuring and evaluation of social situations and to initiate adequate behaviour. An additional special functional role of stereotypes 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, 19). So, there is empirical evidence in support of the person model theory for both levels, non-conceptual and conceptual mindreading.

### 5.3 *What are the advantages of the person-model theory?*

The thesis that we can directly perceive basic mental phenomena avoids the implausible claim central to Theory-Theory that we *always* have to rely on theoretical assumptions and make inferences when we try to understand other minds. Young infants at around one year of age do not seem to rely on any theory even if we presuppose only a basic understanding of what a theory is.<sup>15</sup> We argued that there is a non-conceptual understanding of

---

15. One version of Theory-Theory is the so-called Child-Scientist Theory. Representatives argue that the understanding of other minds starts without an ability to understand false beliefs. This is learned—in a scientific fashion—in the first four years of life. For a critical discussion

other minds. We can also avoid the shortcomings of Simulation-Theory which cannot account for high-level attribution of beliefs and desires which are used as input for a decision making process—a deficit we pointed out in Goldman's theory. Furthermore, Goldman is forced to include Theory-Theory in his hybrid account since he cannot account for our understanding of people with radically different mental dispositions (like people suffering from schizophrenia, autism, Capgras syndrome etc.). In order to capture this, we offer the notion of a conceptual understanding of others by creating, using and improving person images of individuals and groups which allow us to estimate quickly the mental situation of others. Finally, we can avoid the implausible implication of a pure Interaction-Theory that even complex mental states can be directly perceived. Instead, we offer the view that person schemata are essentially based on direct perception while person images are essentially relying on the interpretation of situations involving background knowledge and the construction of narratives. As we have emphasized in our criticism of Goldman's Simulation-Theory, we have to acknowledge various quite different strategies of understanding others. We have distinguished, broadly, non-conceptual from conceptual understanding. Which of these strategies is (or *needs* to be) employed in order to understand another person depends crucially (a) on the person in question and our prior relation to and familiarity with that person (that is, basically, on the richness of our person-model regarding that person), (b) on the situation and context and, finally, (c) on the type and complexity of the mental state(s) in question. All these three dimensions have to be taken into account in developing a persuasive theory of understanding other minds.<sup>16</sup>

---

see (Goldman 2006, Chap. 4). Regarding our view, it is sufficient to say that even according to child-scientist views it is not justified to attribute children a *theory* before they have learned to master the false-belief task. There has recently been some debate about the onset of this competence which we cannot touch in this article.

16. For example, in case we are very close to the person, we may rely on a non-conceptual way of understanding. For example, we rarely need to theorize about what our own children may think or feel because they are very close to us and we have developed a rich person model of them. But we may also rely on this non-conceptual strategy if we observe a complete stranger displaying a very familiar type of behaviour. But in case we see a person for the first time *and* see her displaying a behaviour that is quite strange to us, then we may need to employ various strategies at once in order to understand what she is up to; we may need to consult person images of other people and our own person model of ourselves. Similarly, when our children reach puberty, they may display quite strange types of behaviour such that we may need to theorize about what the kids are up to, despite our rich person model about them.

That being said we wish to clarify the relation between the person model theory and simulation theory: We can account not only for the observation that simulation strategies are sometimes used but also indicate to which extent they are used: The simulation strategy of high-level mindreading as suggested by Goldman (estimating the actual beliefs and desires of someone, introducing those attitudes into my reasoning systems producing a decision and projecting this decision to someone else) is used in situations when I have evidence that another person is psychologically very similar to me: If I believe that the other person is of the same gender, age and in the same professional and private life situation, then I start to understand that person mainly on the basis of simulation. If I discover differences as time goes by, I start to quarantine individual differences and step by step develop an individual person image different from my self-image which becomes the basis of understanding that person. Such cases of strong psychological similarities are rare.<sup>17</sup> Quarantining our own beliefs and desires is not a problem in our theory because once I have developed a person image of Peter, for example, I can always rely on that image to understand Peter and may rely on another person image (or images) if faced with limits in understanding Peter's behaviour in order to improve and adjust my person model of him. But then it is in no way clear that we (sometimes or even always) use our person image of ourselves. If I have evidence that in a certain situation, Peter behaves more like Karl (whose behaviour is also very different from mine), I immediately switch to my person image of Karl, using it as a pattern to understand Peter. Our introductory story of Ralph who believes that Ortcutt is a spy can be nicely accounted for: We all usually acquire a stereotype of a spy by reading crime stories and watching movies. Such a stereotype is used by Ralph to understand Ortcutt's behaviour although he has only very sparse observational information about him.

The person-model theory can also account for the observation that, ontogenetically, human beings gain a better understanding of other people step by step. After first only relying on person schemata, we then develop person images, which become richer and richer.

We can learn to understand a person who is psychologically radically different from us without ever being able to simulate her. We can simply

---

17. If a person does only have a very impoverished self-model then of course s/he can detect people which seem to be like herself more easily. It is implausible that I have to rely on simulation to understand decisions which are common to almost all humans because it is cognitively much more economical to rely simply on our stereotypes of common human behaviour.

enhance our repertoire of person images by acquiring (or adopting) person images, which we find in story telling, literature, sciences and so on. We also acquire person images of familiar persons with detailed knowledge of them even if they are essentially distinct from ourselves concerning their psychological dispositions. Another advance of the person model theory is that it can account for the fact that we sometimes make explicit evaluations of a person that do not fit with our behaviour towards her: If I consciously evaluate a person as trustworthy and friendly but at the same time my nonverbal communication signals that I am suspicious and notice some aggressiveness, then this can be described as a conflict between my person model and my person schema of the other.<sup>18</sup>

Finally, it is an advantage that we offer a theory which can account for the fact that, normally, first-person understanding and third-person understanding are roughly on one level. *Person schemata* are the product of automatic psychological processes which develop and are used to treat first- and third-person-information. To construct complex *person images* we have to learn the classifications of mental and physical dispositions, which are then used in both cases, for myself and for other people. If someone has a strong tendency to use only their own psychological dispositions and mind-set to understand other people then this leads to a strong egocentric bias which puts a limitation on an adequate social interaction. An extreme example of such a bias is manifested in egomania.

To sum up: This alternative view avoids the disadvantages and shortcomings of Theory-Theory, Simulation-Theory, Interaction-Theory, and the Narrative Practice Hypothesis while retaining their benefits. At the same time, it can account for many of our folk-psychological intuitions as well as scientific results in psychology and neuroscience. We are therefore optimistic that this sketch of the person-model theory can be further developed into a full-blown theory.

## REFERENCES

- Baron-Cohen, Simon 1995: *Mindblindness. An Essay on Autism and Theory of Mind*. Cambridge, Mass.: MIT Press.
- Baron-Cohen, Simon, Alan M. Leslie and Uta Frith 1985: "Does the autistic child have a 'theory of mind'?" *Cognition* 21, 37–46.

---

18. In the same line some cases of self-deception can be characterized as cases in which my self-image is different from my self-schema.

- Bente, Gary, Ansgar Feist and Stephen Elder 1996: "Person perception effects of computer simulated male and female head movement". *Journal of Nonverbal Behavior* 20, 213–228.
- Bente Gary, Nicole C. Krämer, Anita Petersen and Jan Peter de Ruiter 2001: "Computer animated movement and person perception. Methodological advances in nonverbal behavior research". *Journal of Nonverbal Behavior* 25(3), 151–166.
- Damasio, Antonio R. 2003: *Looking for Spinoza. Joy, sorrow, and the feeling brain*. London: Heinemann.
- Davies, Martin, Max Coltheart, Robyn Langdon and Nora Breen 2001: "Monothematic delusions: Towards a two-factor account". *Philosophy, Psychiatry, and Psychology* 8, 133–58.
- Dennett, Daniel C. 1971: "Intentional Systems". *The Journal of Philosophy* 68, 87–106.
- 1987: *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- de Vignemont, Frédérique 2009: "Drawing the boundary between low-level and high-level mindreading". *Philosophical Studies* 144(3), 457–466.
- de Vignemont, Frédérique and P. Fournier 2004: "The sense of agency: a philosophical and empirical review of the Who system". *Consciousness and Cognition* 13(1), 1–19.
- Doise, Willem and Anne Sinclair 1973: "The categorization process in intergroup relations". *European Journal of Social Psychology* 3, 145–157.
- Ekman, Paul, E. Richard Sorenson and Wallace V. Friesen 1969: "Pan-cultural elements in facial displays of emotion". *Science* 164, 86–88.
- Fletcher, P., F. Happé, U. Frith, S. C. Baker, R. J. Dolan, R. S. Frackowiak, C. D. Frith 1995: "Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension". *Cognition* 57, 109–128.
- Frey, Siegfried 1999: *Die Macht des Bildes. Der Einfluss der nonverbalen Kommunikation auf Kultur und Politik*. Göttingen: Huber.
- Frith, Uta and Christopher D. Frith 2003: "Development and neurophysiology of mentalizing". *Philosophical transactions of the Royal Society London Series B* 358, 459–473.
- Gallagher, H. L., F. Happé, N. Brunswick, P. C. Fletcher, U. Frith, and C. D. Frith 2000. "Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks". *Neuropsychologia* 38: 11–21.
- Gallagher, Shaun 2001: "The practice of mind: Theory, simulation, or interaction?" *Journal of Consciousness Studies* 8 (5-7), 83–107.
- 2005: *How the body shapes the mind*. Oxford: OUP.
- 2007: "Simulation trouble". *Social Neuroscience* 2/3, 353–365.
- 2008a: "Direct perception in the intersubjective context". *Consciousness and Cognition* 17, 535–543.

- 2008b: “Another look at intentions: A response to Raphael van Riel’s Seeing the invisible”. *Consciousness and Cognition* 17, 553–555.
- Gallagher, Shaun and Andrew N. Meltzoff 1996: “The Earliest Sense of Self and Others: Merleau-Ponty and Recent Developmental Studies”. *Philosophical Psychology* 9, 213–236.
- Gallagher, Shaun and Dan Zahavi 2008: *The phenomenological Mind. An Introduction to Philosophy of Mind and Cognitive Science*. London: Routledge.
- Gallese, Vittorio, Luciano Fadiga, Leonardo Fogassi and Giacomo Rizzolatti 1996: “Action recognition in the premotor cortex”. *Brain* 119, 593–609.
- Georgieff, Nicholas and Marc Jeannerod 1998: “Beyond consciousness of external reality. A ‘Who’ system for consciousness of action and self-consciousness”. *Consciousness & Cognition* 7(3), 465–477.
- Goldman, Alvin I. 1989: “Interpretation Psychologized”. *Mind and Language* 4, 161–185.
- 2006: *Simulating minds. The Philosophy, Psychology, and Neuroscience of Mind-reading*. Oxford: OUP.
- forthcoming: “Mirroring, mindreading and simulation”. To appear in Jaime A. Pineda (ed.), *Mirror Neuron Systems: The Role of Mirroring Processes In Social Cognition*.
- Gopnik, Alison 1993: “How we know our minds: The illusion of first-person knowledge of intentionality”. *Behavioral and Brain Sciences* 16, 1–15, 90–101.
- Gopnik, Alison and Andrew N. Meltzoff 1997: *Words, thoughts, and theories*. Cambridge, Mass.: Bradford, MIT Press.
- Gopnik, Alison and Henry M. Wellman 1994: “The ‘Theory-Theory’”. In: Lawrence A. Hirschfield and Susan A. Gelman (eds.), *Mapping the mind: Domain specificity in culture and cognition*. New York: Cambridge University Press, 257–293.
- Gordon, Robert M. 1986: “Folk Psychology as Simulation”. *Mind and Language* 1, 158–171.
- Heal, Jane 1986: “Replication and Functionalism”. In: Jeremy Butterfield (ed.), *Language, Mind, and Logic*. Cambridge: Cambridge University Press, 135–150.
- Heider, Fritz and Marianne Simmel 1944: “An experimental study of apparent behavior”. *American Journal of Psychology* 57, 243–259.
- Hobson, Peter 2002: *The Cradle of Thought*. London: Macmillan.
- Hutto, Daniel D. 2008: *Folk-psychological narratives*. Cambridge, Mass.: MIT Press.
- Jacob, Pierre and Marc Jeannerod 2003: *Ways of seeing. The scope and limits of visual cognition*. Oxford: OUP.
- Kanwisher, Nancy 2001: “Neural events and perceptual awareness”. *Cognition* 79, 89–113.

- Leslie, Alan M. 1987: "Pretense and Representation: The origins of 'Theory of Mind'". *Psychological Review* 94, 412–426.
- Macrae, C. Neil and Galen V. Bodenhausen 2000: "Social cognition: Thinking categorically about others". *Annual Review of Psychology* 51, 93–120.
- Meltzoff, Andrew N. and Jean Decety 2003: "What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience". *Philosophical transactions of the Royal Society London Series B* 358, 491–500.
- Meltzoff, Andrew N. and M. Keith Moore 1977: "Imitation of facial and manual gestures by human neonates". *Science* 198, 75–78.
- Moore, D. G., R. P. Hobson and A. Lee 1997: "Components of person perception: An investigation with autistic, non-autistic retarded and typically developing children and adolescents". *British Journal of Developmental Psychology* 15, 401–423.
- Newen, Albert and Andreas Bartels 2007: "Animal Minds: The Possession of Concepts". *Philosophical Psychology* 20/3: 283–308.
- Newen, Albert and Kai Vogele 2003: "Self-Representation: The Neural Signature of Self-Consciousness". *Consciousness & Cognition* 12, 529–543.
- Newen, Albert and Gottfried Vosgerau 2007: "A representational theory of self-knowledge". *Erkenntnis* 67, 337–353.
- Nichols, Shaun and Stephen P. Stich 2003: *Mindreading. An integrated account of pretence, self-awareness and understanding other minds*. Oxford: OUP.
- Oakes, Penelope J., S. Alexander Haslam and John C. Turner 1994: *Stereotyping and social reality*. Malden, MA: Blackwell.
- Olsson, Andreas and Kevin N. Ochsner 2007: "The role of social cognition in emotion". *Trends in cognitive sciences* 12(2), 65–71.
- Pauen, Sabina 2000: "Wie werden Kinder 'Selbst'-Bewusst? Entwicklung in früher Kindheit". In: Albert Newen and Kai Vogele (eds.), *Selbst und Gehirn. Menschliches Selbstbewusstsein und seine neurobiologischen Grundlagen*. Paderborn: Mentis, 291–312.
- Ratcliffe, Matthew J. 2007: *Rethinking Commonsense Psychology: A Critique of Folk Psychology, Theory of Mind and Simulation*. Basingstoke: Palgrave Macmillan.
- Reddy, Vasudevi 2008: *How infants know minds*. Cambridge, Mass.: Harvard University.
- Rizzolatti, Giacomo and Laila Craighero 2004: "The mirror neuron system". *Annu. Rev. Neurosci.* 27, 169–192.
- Rizzolatti, Giacomo, Luciano Fadiga, Vittorio Gallese and Leonardo Fogassi 1996: "Premotor cortex and the recognition of motor actions". *Cogn. Brain Res.* 3, 131–141.

- Santos, Natacha S., Nicole David, Gary Bente and Kai Vogeley 2008: "Parametric induction of animacy experience". *Consciousness and Cognition* 17(2), 425–37.
- Saxe, R. and Nancy Kanwisher 2003: "People thinking about people: The role of the temporo-parietal junction in 'theory of mind'". *NeuroImage* 19, 1835–1842.
- Van Riel, Raphael 2008: "On how we perceive the social world. Criticizing Gallagher's view on *direct perception* and outlining an alternative". *Consciousness and Cognition* 17(2), 544–552.
- Vogeley, K., P. Bussfeld, A. Newen, S. Herrmann, F. Happé, P. Falkai, W. Maier, N.J. Shah, G.R. Fink and K. Zilles 2001: "Mind reading: Neural mechanisms of theory of mind and self-perspective". *Neuroimage* 14, 170–181.
- Vogeley, Kai and Albert Newen 2002: "Mirror Neurons and the Self Construct". In: Maxim Stamenov and Vittorio Gallese (eds.), *Mirror Neurons and the evolution of brain and language*. Amsterdam and Philadelphia: John Benjamins Publishers, 135–150.
- Volz, Kirsten G. 2008: „Ene mene mu—insider und outsider“. In: Ricarda Schubotz (ed.), *Other minds. Die Gedanken und Gefühle anderer*. Paderborn: Mentis, 19–30.
- Volz, Kirsten G., Thomas Kessler and D. Yves von Cramon (under review): "In-group as part of the self: In-group favoritism is mediated by medial prefrontal cortex activation." *Social Neuroscience*.
- Vosgerau, Gottfried and Albert Newen 2007: "Thoughts, motor actions and the self". *Mind and Language* 22(1), 22–43.
- Wicker, Bruno, Christian Keysers, Jane Plailly, Jean-Pierre Royet, Vittorio Gallese and Giacomo Rizzolatti 2003: "Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust". *Neuron* 40, 655–664.
- Wimmer, Heinz and Josef Perner 1983: "Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception". *Cognition* 13, 103–128.
- Zinck, Alexandra and Albert Newen 2008: "Classifying Emotion: A Developmental Account". *Synthese* 162(1), 1–25.