

THEORIES OF UNDERSTANDING OTHERS: THE NEED FOR A NEW ACCOUNT AND THE GUIDING ROLE OF THE PERSON MODEL THEORY

Abstract. *What would be an adequate theory of social understanding? In the last decade, the philosophical debate has focused on Theory Theory, Simulation Theory and Interaction Theory as the three possible candidates. In the following, we look carefully at each of these and describe its main advantages and disadvantages. Based on this critical analysis, we formulate the need for a new account of social understanding. We propose the Person Model Theory as an independent new account which has greater explanatory power compared to the existing theories.*

1. Introduction

Humans are hyper-social beings that are highly dependent on adequate interaction with others. Right after birth our survival depends on social interaction, and this remains a key aspect of biological, economic and social success throughout our entire lives. Given the phylogenetic and ontogenetic relevance of social interaction, researchers across disciplines aim for an adequate theory of how humans are able to understand others. This aim has not yet been definitively reached, but recently the debate has received new input after a decade of stagnation (in the 90s) within a philosophical debate virtually restricted to the choice between Simulation Theory and Theory Theory. Now, multiple accounts are on the table. This motivates us to clarify the main positions, their arguments and their relations to each other. We proceed as follows: first, we dedicate one section apiece to each central (class of) positions, namely Theory Theory, Simulation Theory and Interaction Theory. Based on the advantages and disadvantages that these accounts reveal, we argue in a second step in favor of what we call Person Model Theory.

1 For fruitful contributions we would like to thank Atesh Koul and Julia Reh as well as two anonymous reviewers. The paper was supported by the DFG-Research Training Group (no. GRK2185) "Situated Cognition" and by the DFG-project (NE 576/14-1) "The structure and development of understanding actions and reasons".

2. Philosophical Theories of Mind-Reading

Three competing alternatives enjoy the lion's share of discussion within the philosophy of 'mind-reading'. Theory Theory introduces theory-based inferences relying on folk-psychological rules as the central basis for understanding others. This view may account for a great deal of explicit and reflective social understanding after age 4 or 5. However, it seems to be inadequate to account for the intuitive understanding of others which develops rather early in infancy and remains active even after we have acquired the capacity for explicit theory-based inferences. Simulation Theory is especially suited to account for the early intuitive social understanding that takes place on the basis of simulating the mental state of other subjects. However, one main deficit is that the simulation process may not be possible in many real-life situations, including observations of persons with mental disorders or radically different cultural backgrounds. Moreover, it appears questionable whether this theory is an adequate framework to integrate the basic mirror neuron story on which most defenders of this account strongly rely. Interaction Theory argues plausibly for the importance of direct and smart perception in social understanding as well as for the distinguishing role of online interaction. However, when seen in the light of recently available empirical evidence, Interaction Theory overestimates the primacy of basic forms of direct social coordination.

2.1 *Theory Theory*

The core idea of "Theory Theory" (TT) is the claim that the capacity to understand others is based on a folk-psychological theory that is used for systematic inferences. Humans employ a folk-psychological theory (i.e. an abstract and coherent system of law-like assumptions) to derive the mental states of others, such as their beliefs, attitudes, desires or emotions, and to thereby anticipate their future behavior.

A modular version of TT is provided by Baron-Cohen (1995), according to which the human psychological system is composed of various modules which interpret the world in accordance with an inborn organizational structure evolved through natural selection. Each module is tailored to solve a certain adaptive problem, and correspondingly there exists a specialized mindreading system designed to comprehend and predict the behavior of other subjects. Due to its linkage with perceptual processes, this system enables the recognition of visible cues reliably indicating the internal mental states of others. The flexible and fast inference of the complete range of mental states from observable signals depends on the management of the 'Theory-of-Mind-Mechanism' (ToMM).

In contrast, Gopnik & Wellman (1992) assume that children gradually develop a Theory-of-Mind (ToM) ability that is based on the same cognitive

mechanisms that adults apply in the development of scientific theories. During a constant learning process, children generate general assumptions about unobservable entities, formulate expectations, adjust their theory in accordance with the evidential data they collect through experience, invent auxiliary hypothesis and replace their theoretical constructs and rules in the light of continuously occurring counter-evidences. The progressive improvement of this causal-explanatory theory about how others come to perform specific actions finally leads to a coherent representational system of propositional attitudes.

Accordingly, the main difference between those two accounts concerns the acquisition of the ability to explicitly represent the mental states of others. Gopnik & Wellman claim that the ToM ability is based on a psychological theory which passes through the same dynamic process of prediction, falsification and adjustment as do scientific theories, until it reaches the status of a mature theory which complies with the demands of a full-blown ToM ability. Conversely, Baron-Cohen argues for a phased maturing of distinct innate modules, where the ToMM comes into operation only in the final stage. Nevertheless, both accounts agree on the assumption that from a certain point in infantile development humans refer to a complex theory-like structure of mentalistic knowledge to infer the propositional attitudes of others.

This idea is supported by experimental studies proving that approximately around the age of 4 children possess the mentalistic abilities to pass the so-called 'false belief task', an experimental setup which has been implemented in widely known versions by Wimmer & Perner (1983). According to Baron-Cohen (1995), at this age children have mastered the use of the ToMM, while Gopnik & Wellman (1992) interprets the results as a further improvement of their causal-explanatory theory based on previous experiences of false prediction. Independently from the question of ontogenetic genesis, the employment of theory-based inferences is treated as the general epistemic strategy used by older children and adults in everyday life. This inferential mindreading mechanism proves especially useful in understanding other subjects which differ fundamentally in their mindset and their behavioral patterns from oneself. Such situations might occur in contact with members of other cultures, with persons suffering from mental diseases, or with animals (Newen 2015a). Furthermore, humans tend to use the epistemic strategy of TT when a social situation offers merely a small number of perceptual cues for the internal state of persons involved (Baron-Cohen 1995).

According to TT, theory-based inferences are the primary mentalistic strategy. However, TT overlooks the possibility of having direct access to many basic mental phenomena, simply by simulating the other's situation or directly perceiving their mind state. It overintellectualizes intuitive understanding in early infancy and it underestimates the role of second-person involvement as well as the role of one's own experiences in understanding others.

The TT assumes the employment of a third-person viewpoint towards another person's mental states in a manner analogous to scientific inquiry. Nevertheless, the ability to comprehend and predict behavior in mentalistic terms becomes particularly important in interpersonal cooperation where the mindreading person does not merely function as an observer but as a dynamic interacting part (Di Paolo & De Jaegher 2012). The worry is that the observational stance which is usually adequate in science is only an exceptional perspective in understanding others, while humans are frequently involved in second-person interactions (Vogeley, Schilbach & Newen 2013; Schilbach et al. 2013). Furthermore, instead of relying on a theory, people sometimes just rely on their own sparse experiences in similar situations as a basis for mindreading (Goldman 1992).

The method of the TT is grounded on the assumption that internal mental states are only accessible due to complex inferences whereas observational behavior constitutes the evidential basis for further theoretical considerations. Conversely, understanding others does not always require such intellectual capacities (Gallagher 2008). At least within human culture, we seem to possess the universal ability to directly perceive basic emotions (Ekman & Friesen 1971; Gallagher 2008; Newen et al. 2015). Even young infants are able to intuitively understand others although they have not yet acquired an explicit or implicit theory of systematically interconnected beliefs. Ontogenetic studies clearly demonstrate that infants of less than one year of age are sensitive in their reactions to the affective expressions of caretakers, as in the visual cliff experiment (Sorce et. al. 1985), and they expect a smooth interaction pattern which leads to irritation if not used, as in the still face paradigm (Weinberg et. al. 2008).

2.2 Simulation Theory

In contrast to TT, 'Simulation Theory' (ST) dismisses the assumption that humans use a specific 'theory' to understand other people's minds. Rather, subjects simulate the others' situations and 'put themselves in the other's shoes'. The ST account does not need to presuppose a generalized set of laws similar to science, and it is characterized as information-poor mindreading (Goldman 1992) as it does not presuppose an interconnected set of beliefs or belief-like information. However, as argued by Gordon (1986, 1992), simulating other minds does not merely mean to project one's own situation, as it also requires necessary adjustments concerning other persons and their perspectives. It is suggested that ST is routed in phylogenetically and ontogenetically basic mechanisms, taking advantage, for instance, of human abilities to read gaze direction or to imitate others (Gallese & Goldman 1998). Simulation enables subjects to generate explanations for the behavior of others and to predict how they are most likely to act in the future (Goldman 1989; Gordon 1992; Spaulding 2010). Some of these simulations

are thought of as conscious and voluntary processes (Goldman 2006), others as unconscious and automatic in the sense that they do not require access or control over the stimulation processes (Gordon 1992).

Egocentric errors or biases, i.e. the influences of the mindreader's own mental states on the ascription of mental states to others, provide the first evidence for ST (Goldman & Jordan 2013). A paradigmatic example is the so-called curse of knowledge. This term designates the phenomenon that the participant's own knowledge influences his attribution to another person, although the participant is informed about the difference between their own and the other person's knowledge (Birch & Bloom 2003; Camerer, Loewenstein & Weber 1989; Nickerson 1999). ST also gathers support from neurophysiological studies where, for instance, amygdala lesions do not only strongly reduce the experience of fear in patients, but also their ability to recognize the fear of others on the basis of their facial expressions (Adolphs et al. 1994).

The interdependence of first-person experience and third-person observation receives further evidence from the discovery of mirror neurons. The class of mirror neurons, first discovered in monkeys, is active both during the performance of an action as well as during the observation of another individual performing this very action (Di Pellegrino et al. 1992; Rizzolatti & Craighero 2004). It has been proposed that when we observe someone perform an action, activation in our mirror neuron system simulates the action 'as if' we were performing it. The discovery of mirror neurons is supposed to be the most striking evidence for ST, as they are supposed to constitute the neural realization of at least the automatic forms of simulation. The mirror neuron system has thus been proposed as the basis for our understanding of others (Gallese & Goldman 1998; Keysers & Gazzola 2009; Rizzolatti & Craighero 2004; Sinigaglia & Rizzolatti 2008).

Despite the supporting evidences for ST, the theory faces several main issues. First, one can think of many cases in which subjects reliably predict the experience and behavior of others without being able to simulate them. For instance, ST is not necessary to understand persons with mental disorders, such as patients suffering from delusion of persecution, or persons who exhibit idiosyncratic, irrational behavior (Tversky & Kahneman 1974). The same also holds for beings involved in radically different cultures (Newen & Schlicht 2009). Their minds are simply too different from one's own to apply the epistemic strategy of simulation. Nonetheless, persons who possess general or specific knowledge about the respective subjects are able to understand what is going on in their minds and to successfully interact with them (Newen 2015a). This general or specific information we make use of can be learned as rules-of-thumb or an explicit theory, e.g. how to deal with a schizophrenic family member, without being able to simulate this person. Understanding based on behavioral rules-of-thumb or a theory can

be quite advanced and enables smooth interaction despite lacking short-term or even long-term simulating or intuitive access to the deviating mindset of the others. To justify this relevance of rules-of-thumb or explicit theories in cases of mental disorders, we appeal to everyday experiences in dealing with persons with different mindsets (and without knowing this mindset), but we can also rely on studies with Asperger autistic people: they have a large deficit in all types of intuitive or simulative epistemic access to others but they can learn to partially compensate by learning to apply explicit theoretical rules. This indicates that simulation is not necessary and that non-autistic people rely on a plurality of epistemic strategies, not only simulation which can be lacking (for details see section 3.2 the pathology argument).

Second, ST remains limited in the sense that it adopts a first-person perspective in which the simulating individual is still considered as an observer (Gallagher 2008). Our requirements for understanding others' actions, however, is critical when we are interacting with them (Schilbach et al. 2013) whereby this online interaction is often realized with non-simulative but complementary actions (de Bruin et al. 2012). Third, the discovery of mirror neurons does not so far explain the relation between the first-person and third-person perspective. Mirror-neurons encode for certain types of actions and emotion, but they do not provide an answer to the question of how we attribute internal states to others on the basis of these neural processes. Moreover, the neural correlate in the case of third-person attribution of, for instance, beliefs does not involve the most characteristic correlates of first-person attribution of belief (Vogeley et al. 2001; Vogeley & Newen 2002), whereas ST would expect such an involvement. Thus, despite the important discovery of the mirror neuron system, its role in the process of understanding others still needs to be worked out in detail and its function in cases of simulation (which might sometimes happen) needs to be complemented by further neural processes. As long as this part of the story is missing, the mirror neuron system remains an interesting and still important component for automatic social processing (Neufeld et al. 2016), but this component still needs to be integrated into a theory of understanding others.

2.3 Interaction Theory

Interaction Theory (IT) is a phenomenologically inspired approach which claims that we understand others primarily and most importantly in situations of direct social interaction, which leads to the distinction between *online* and *offline* forms of social understanding (Frith & Frith 2003). More precisely, IT combines at least two claims: one about the important role of direct perception of mental states of others independent from any inferences (Gallagher 2008), and one about the primacy of understanding by adequate interaction (Hutto & Gallagher 2008).

IT characterizes human ‘mindreading’ as a form of smart perception. This means that the content of our perceptual experience can be rich and include mental phenomena in the sense that we can directly perceive the internal states of other subjects. While some argue that the contents of perceptual experiences are exclusively low-level properties (Tye 1995), in recent years many people have argued that the contents of perceptual experiences can also involve high-level entities such as causal relations (Butterfill 2009; Siegel 2009), actions and agency (Gao et al. 2009; Rutherford & Kuhlmeier 2013). In the same way, a phenomenological perspective is often used to argue for the rich content of our perceptual experience in social cognition, prominently defended by Gallagher (2008) and Zahavi (2011). The general line of argument can be roughly characterized as follows: perceptual experiences can be cognitively penetrated and they can thereby involve a rich content (McPherson 2012; Vetter & Newen 2014; Newen & Vetter 2016). Expert perception, we may say, is different from the perception of laypersons. A chess expert has a richer perceptual content when looking at a chessboard compared to a novice (Newen 2017). Since humans are hyper-social beings and, thus, most likely experts in understanding others, we are able to have a rich content in our social perception, e.g. in the perception of others’ emotions (Zahavi 2011; Marchi & Newen 2015) or intentions (Pacherie 2005).

The relevance of direct perception has been convincingly argued for, and as a consequence even some representatives of TT have recently started to include direct perception as an important epistemic tool (Herschbach 2012; Carruthers 2015). Thus, direct perception appears as a certain kind of epistemic strategy that might be employed in different forms of mind-reading, even from a third-person perspective. In contrast, IT accounts rely in large parts on the assumption that the central constituent of understanding others is direct perception in online interaction which highlights the relevance of the second-person perspective in mind-reading (Gallagher 2002, 2008). Different versions of IT allow for several strategies of understanding others, all of which assume the primacy of understanding by interaction (Hutto & Gallagher 2008). De Jaegher & Di Paolo (2007), for instance, claim that the constitutive feature of all cases of online interaction is participatory sense-making where this is explained in terms of coordination. According to them, the process of coordination in interaction is constitutive in many cases of social understanding. The main examples to support this claim are cases of special joint action based on mutual social understanding, such as ballroom dancing.

Some evidence for the relevance of social interaction for social understanding is drawn from developmental psychology, which distinguishes the capacity for primary, secondary and sometimes in addition tertiary intersubjectivity (Trevarthen & Hubley 1978; Trevarthen 1979). Primary intersubjectivity involves the ability to reciprocate in face-to-face exchange and starts from two months of age onwards and thereby goes beyond the very early pure imitation abilities. It is, for instance, demonstrated in the still-face-

paradigm (Bertin & Striano 2006). Consequently, we have a communication basis that allows even infants to exchange and read common cues via bodily movements, gestures, facial expressions, eye direction, etc. Secondary intersubjectivity is typically realized when triadic intentional communication begins, e.g. in interactions involving joint attention which start at approximately 9 months. This secondary level involves the understanding of other people while acting together in a pragmatic context. It permits sharing and coordinating with another person's attention, feelings and intentions toward a third object, event, or action (Trevarthen & Hubley 1978). While it is assumed that primary intersubjectivity is innate and allows even newborns to perceive other person's mental processes, secondary intersubjectivity develops later in the first year of life. Tertiary intersubjectivity develops when children aged 4 begin to employ an ethical stance by beginning to manifest explicit rationale about what is right and wrong, as well as explicit attitudes about others' mental states (Trevarthen 2006).

In addition, there is now more and more evidence that social cognition is fundamentally different when we are in interaction with others rather than merely observing them. This is shown by systematic investigations of the underlying neural processing, e.g. in a test of observing facial expressions which are either directed towards oneself or towards another. While self-directed facial expressions lead to a differential increase of neural activity in the ventral portion of the medial prefrontal cortex and the (superficial) amygdala, other-directed facial expressions result in a differential recruitment of medial and lateral parietal cortex (Schilbach et al. 2006). In another study (Schilbach et al. 2010) of two persons either realizing joint attention towards an object or looking at different objects, it was shown that joint attention had a specific neural profile which closely matches with the so-called mentalizing network relying on the medial prefrontal cortex and posterior cingulate cortex. Furthermore, it was shown that producing joint attention (e.g. directing someone else's gaze toward an object) activated the ventral striatum (i.e. reward system). This indicates that activating joint attention is pleasurable for healthy people. Together with other evidence, this triggered the claim that we should presuppose a *second-person neuroscience* (Schilbach et al. 2013) and, thus, it is convincingly argued that understanding others in a situation of online-understanding is systematically different from understanding others by observation without interaction.

The most important insight is delivered by the claims that online-understanding is a specific form of understanding in contrast to offline-understanding, and that direct perception plays a decisive role in social understanding. This is the strongest feature of IT, but it still leaves us with the open question whether online understanding is in fact primary to offline understanding, and if so in which sense – phylogenetically, ontogenetically or even constitutively. It remains questionable to what extent observations of simple coordination can be generalized to all cases of social understanding

and whether it is prior in comparison to the diverse other forms of social understanding (Andrews 2012; Newen 2015a). The evidence here is uncertain but evolutionary considerations may speak for the claim that both are equally relevant strategies of understanding. To survive as social beings, humans need to learn from both interaction and observation as much and as soon as possible. Thus, a primacy claim leaves the burden of proof on the side of IT.

Furthermore, IT overlooks the importance of the long-term social relationships which are habitual and re-activated in social interactions, e.g. in the case of understanding a familiar person. This long-term person-centred information can become strongly relevant in shaping an online interaction, much more so than any specific information about the situation in which one deals with this person, and it is also relevant in offline understanding, such as when trying to understand the familiar person while discussing him or her with a friend. This criticism can be condensed into one core difference for which IT cannot account: namely, the difference between the social understanding of a person's actions in one and the same situation type, where in one case the person is a complete stranger and in the other a well-known person such as a family member or a friend. This is especially relevant since this difference is already implemented in early infancy, e.g. the phenomenon of infant shyness in which infants react shyly to adult strangers, which manifests during the third quarter of the first year.² The relevance of prior information in the evaluation of a person's mind-set is also reflected in empirical studies investigating the impact of stereotypes. Culturally anchored stereotypes (Macrae & Bodenhausen 2000) and stereotypes in general (Macrae & Quadflieg 2010) substantially shape our understanding of others (review: Newen 2015a, sec. 5.1–5.3).

3. Person Model Theory

3.1. Definition

The central idea of the 'Person Model Theory' (PMT) is twofold: On the one hand, we need to accept that humans use *a multiplicity of epistemic strategies* (theory-based inferences, simulation, direct perception, contextual or narrative embedding) to account for all cases of understanding others. On the other hand, we need to take into account that humans rely on prior information stored in form of *person models* and *situation models*. As such,

2 Defenders of IT may reply that they can include *memorized* interactions to account for these facts. However, this would require substantial alterations of the innate proposal of IT in accepting memorized models of other persons. Another move would be to claim that the memorized information is available in the form of narratives, since those are an additional tool in IT. Although narratives are an important instrument to enrich information about others which unfolds from 2 years of age onwards (Hutto 2008; Newen 2015), they cannot account for the relevant sensitivity in early infancy.

the PMT according to Newen (2015a, 2017) aims to answer two mainly independent questions.³

The first question asks which epistemic strategy humans use to access the mental states of others and to gather information about them. Concerning the *epistemic strategy*, PMT defends the multiplicity view: we do not rely on one epistemic strategy as is suggested by most proposals in the literature (e.g. ST claims that simulation is the only or at least the absolute dominant strategy). On the contrary, human social understanding rather relies on a multiplicity of strategies which are for the most part implicitly activated by contextual cues. These strategies include at least simulation strategies, theory-based inferences, and direct perception, as well as understanding based on social interaction and narratives. A plurality of social understanding was described by Andrews (2012), but she did not work out the important difference between epistemic strategies and the relevant background information which allows a systematic analysis of the rich and varying phenomena of so-called mindreading.

The second question asks how the information we obtain to understand others is stored and organized. The central claim is that information about other humans as individuals or types of persons is stored and organized in *person models*. These models are realized on two levels, namely the implicit level of person schemata and the explicit level of person images. Person models are representational structures like objects files unifying the information about an entity, e.g. another individual or a group of individuals, in a form that is less demanding than a full-fledged theory as proposed in TT.

Concisely, a model contains a unified body of information. This minimal integrated package of information enables us to understand a part of the world, e.g. by enabling us to represent an entity, such as an object, a property, a process, etc. The resulting model enables a person to represent such an entity in our world. If the information of a model is enriched over time, it unfolds into an understanding of the represented entity. In the case of understanding others, the relevant models are especially models of persons. Thus, a person model of an individual constitutes a unified body of information about the relevant individual. A person model typically has a label, namely the person's name, and it is under normal circumstances causally anchored in an entity which is ideally identical to the person in question.

3 The notion of 'epistemic strategies' is understood in a wide sense and is here used equivalent to 'cognitive strategies', i.e. it does not imply specific high-level epistemic abilities like conscious deliberation and, thus, is not necessarily demanding. The answer to the question concerning which epistemic strategy humans use leaves still quite some room for an answer to the second question. One could in principle defend a simulation theory concerning the epistemic strategy and argue that the relevant background information for simulation is organized in person models or organized as a folk psychological theory, etc. Our view consists in the combination of the multiplicity claim concerning epistemic access and the organization of background information as person models.

(Under certain circumstances, this condition might not be fulfilled, either because something went wrong or because the respective person is non-existent, such as in the case of fictional characters). It is further argued that philosophical theories so far have tended to ignore the fact that we usually understand others by relying on rich background information concerning them and their situation. (A possible exception within the representatives of IT is Gallagher (2011).) In addition to person models we also need situation models – as we will argue below.

The two central aspects of PMT, the multiplicity of epistemic strategies and the organization of relevant background information in form of models, are explicated and motivated in more detail below.

3.2 Main Concepts

A Multiplicity of Strategies for Understanding Others

There are two main arguments employed to defend the multiplicity view concerning epistemic strategies. (i) The *ontogenetic* argument indicates that the ontogenetic development of understanding others can best be explained by describing the development of a multiplicity of epistemic strategies such that no strategy is eliminated once acquired. (ii) The *pathology* argument turns on the observation that some cases of mental disorder can best be described by demonstrating that some epistemic strategies are lacking and, thus, others – which are still available – are used as substitutions, even though they often cannot compensate for the complete lack of the original strategies.

Ad (i) The ontogenetic argument: Quite early on, babies rely on *online understanding by coordinated interaction*. They develop an expectation of an interaction scheme as demonstrated by the still-face paradigm. *Direct perception* is very relevant starting from early infancy, as proven by face-based sensitivity for and recognition of emotions based on direct perception (Zahavi 2011; Newen et al. 2015). During ontogeny, we develop further important strategies for understanding others, which also include strategies of offline understanding. It will also be indicated that we cannot observe any general dominance of one of these strategies, but that the activation of a specific strategy is dependent on the context while strategies are often activated in combination. At the age of 9 to 12 months children learn to understand others as participating in *joint attention* and *joint action* (Tomasello 1999), where the latter is demonstrated e.g. by understanding the other as following a plan like jointly constructing a Lego house (at 18 months). At 2.5 years children become sensitive to rules and norms such that they insist that group members follow rules. This involves an *understanding of others as rule-followers*, i.e. as members of the group governed by expectations concerning rule-following behaviour in relevant situations (Rakoczy et al.

2008). Furthermore, there is the well-studied ability to *understand others by explicit false beliefs* (age 4 onwards) which enables explicit theory-based inferences or explicit simulation *strategies* to understand others.⁴ This is correlated with early moral understanding (see above). Finally, *understanding by explicit second-order false beliefs* develops between age 7 to 9 (Wimmer & Perner 1983). Additional epistemic strategies can be fruitfully distinguished as developing later in the process of growing up.

There is consensus that these abilities come gradually, and that abilities acquired early remain intact and in use even when more sophisticated abilities are available. To illustrate: looking at the face of a person, I may directly perceive an expression of anger. However, when I am informed that she is suffering from Parkinson's disease and therefore has severe limitations in controlling her facial expression, I will evaluate the same facial expression quite differently. Despite having a standard 'reading' of emotions from facial expressions, this new knowledge about Parkinson's disease helps me to override my spontaneous perception. Although the direct perceptual impression is still in place, I will override it in this context and use a theory-based inference to reach a new evaluation of the person, relying on other cues including the person's linguistic utterances. This also illustrates the context-dependence of the preference of one strategy over the other.

Ad (ii) The pathology argument: Some mental disorders essentially involve significant deficits in social understanding, and these cases can best be explained such that at least one strategy of the normal bundle of strategies is lacking. This can be illustrated by looking at people with Asperger's syndrome who lack an intuitive understanding of others. They are unable to directly perceive emotions based on facial expressions and they tend to avoid social interactions (Vogeley 2012). Thus, intuitive understanding by primary interaction or direct perception is (almost) unavailable for them. Since they also tend to experience themselves as being different (Vogeley 2012), they do not use simulation as a strategy. Consequently, they can only refer to theory-based inferences that might prove useful to understand others in certain situations (Kuzmanovic et al. 2011). However, they lack an intuitive generalization of this knowledge. Thus, in new or slightly modified situations, they again feel lost since they do not even have a theory on which basis to apply theory-based inferences. Since we have to deal with new or modified 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 availability of multiple strategies of

4 It is an open debate how exactly theory-of-mind abilities and understanding by narratives are related to each other. While Hutto (2008) claims that the latter is more primitive than the former, we presuppose here that *understanding* by narratives is based on a theory-of-mind ability and enriches it. Thus, we do not discuss its role in addition to theory-of-mind abilities.

understanding, they are left mainly with theory-based inferences and need an explicit corpus of knowledge to apply them (since they lack intuitive generalization) (for further arguments concerning the multiplicity view, see Newen 2015a, 2015b; Fiebich & Coltheart 2015; Fiebich 2015).

In sum, social understanding usually relies on a multiplicity of epistemic strategies which are selected in a highly context-dependent manner (as demonstrated with the Parkinson case). Concerning the epistemic strategies of social understanding, we may indicate that social understanding is strongly dependent on the actual context.⁵

Person Models as Unified Information Structures (Person Files)

Having argued for the multiplicity view of epistemic strategies for social understanding, we shall focus now on the organization of the relevant background information in the form of so-called *person models*. There are only a few authors who have considered and developed an account discussed under the label '*Model Theory*' (Newen & Schlicht 2009; Maibom 2009; Godfrey-Smith 2005). The early motivation of Godfrey-Smith (2005) and Maibom (2009) offers a general answer to the status of our folk-psychological knowledge, both adopting the perspective of philosophy of science. Maibom defends a version of the claim that folk psychological knowledge has the status of a model, while understanding a model as a special version of a theory such that she remains in the camp of TT. One important advance is that she argues that a model can be based on *ordinary everyday* knowledge and need not presuppose *special* knowledge. Godfrey-Smith agrees with the latter characterization but also makes important additions by suggesting a specific understanding of 'model' which is different from Maibom's version. According to Godfrey-Smith, a model should not be understood in the tradition of a semantic view of theories. Furthermore, a model can be used in different ways such that we should distinguish between a model and its specific interpretation which he calls a 'construal'. Newen's⁶ account (2015a; 2018) shares the denial of a semantic understanding of 'model' and in addition denies that a model needs to have the structure of a theory. We need a widened understanding of 'model' because this is necessary to enable us to account for the ontogeny of social understanding which is not the focus of either Maibom or Godfrey-Smith.

5 PMT is a full-blown theory of understanding others and is has been developed in a sequence of articles (Newen/Schlicht 2009; Newen 2015a; Newen 2015b; Newen 2018). In this article, the structural organization of background knowledge in form of person models and situation models as well as their interaction is in focus.

6 The use of the third person here indicates that the second author, Coninx, although accepting the general line of the PMT, does not accept all facets of the person model theory as presented by Newen (2015).

In line with Godfrey-Smith, Newen argues that a model can be much more parsimoniously and flexibly used, for instance by relying merely on particular parts of the model to understand certain parts of our world. The used model is not necessarily a theory of the world. Due to its ontogenetic perspective, Newen's (2015a) view on models needs to be distinguished from the claim that "one person predicts another by using a *theoretical model*" (Godfrey-Smith 2005, p. 7). Social understanding involving models can take place at different levels (see previous section) and one basic type is *online understanding by coordinated interaction*. In these cases, the model is not theoretical but perhaps just a memorized interaction schema associated with and expected in relation to a certain person. Models can be rather parsimonious information units which especially cluster information about one person (or one situation).⁷ Focusing on information units about persons, their usage in early ontogeny can be without a theoretical stance: modelling a part of the world can be a different epistemic business than building a theory about it – and in early infancy, it clearly is – while models may unfold into theories during the systematic enrichment and restructuring of information.

Newen's paradigm case of a structure of a model is what is described as a mental file (Perry 1990; Recanati 2012; Newen & Marchi 2016). We can create a mental file of an object with very little information about it and start to systematically enrich and restructure the information unified in this file until it deserves to be called a concept (Newen & Marchi 2016). Since combinations of concepts constitute beliefs and combinations of beliefs constitute theories, there is a cognitive route from parsimonious models of single entities in the world to a theory about complex parts of the world. Thus, folk-psychological knowledge which is quite different and variable in structure is usually given as a model, and may unfold into a theory.

There is for the time being only one philosophical approach which aims to unfold the rather general framework of relying on 'models' into a detailed account of understanding others. This is the recent work of Newen (Newen & Schlicht 2009; Newen & Vogeley 2011; Newen 2014, 2015a, 2018). The central claim here is that relevant information about other humans as individuals or types of persons is stored and organized in person models which are either implicitly available *person schemata* or explicitly available

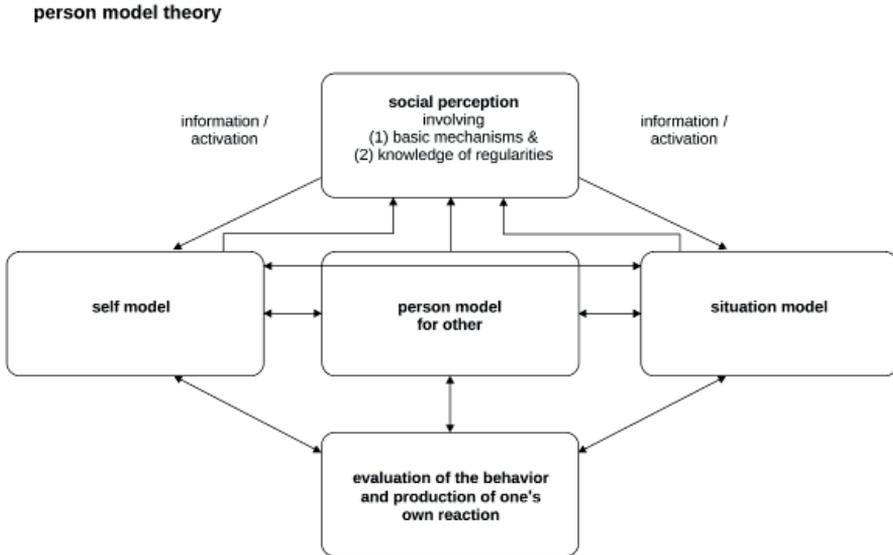
7 Our notion of 'model' can be negatively characterized as different from semantic models which have a very constrained structure as well as from complex models in philosophy of science which always have the status of explicitly available structures. Models positively characterized are systematic informational units which integrate information about an entity into a file which is stored in our memory system. The whole integrated information of the file or a part it can be activated (together with at least one epistemic strategy) and remain implicit to register matching properties, to trigger expectations or evaluations concerning the relevant entity: this information could (but need not) also enter an *explicit* prediction or evaluation of the relevant entity.

person images. The implicit person schema can typically be described as a unity of sensory-motor abilities and basic mental phenomena associated with one human being (or a group of humans). The schema typically functions without any explicit considerations and is activated when directly seeing or interacting with another person. In contrast to this implicit dimension, a person image is constituted by explicitly (i.e. typically consciously) available information concerning physical and mental phenomena associated with and unified to belong to one human being (or a group of humans). 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., anthropologists, students, medical doctors, lawyers, etc. Furthermore, person models are not only created for other people, but also for oneself.

There is recent empirical evidence from neuroscience that we actually construct and rely on person models (Hassabis et al. 2013, see Newen 2015a, section 5.3). It has been shown that there are neural correlates of imagining two central features of the 'big five' in personality psychology, i.e. 'agreeable' in contrast to anti-social personalities, and 'introvert' in contrast to extrovert personalities. Furthermore, it was shown that the combinations of personality types like 'agreeable-ness' and 'extroversion' are represented in a systematic modulation of the medial prefrontal cortex.

Situation models and their intertwinement with person models

An account of full-blown PMT must mention one further component, namely situation models. Humans have the ability to understand others by completely abstracting from the individual: e.g. it can be sufficient to predict the behaviour of a restaurant guest that we expect her to act according to the conventions of a high-class restaurant. This type of understanding is developed together with rule-based understanding of others at the age of 2.5 years (see above). In new contexts, especially in new cultural contexts, we begin by employing an understanding of others mainly on the basis of noticing rule-based behaviour which we discern as being adequate in a situation. For instance, as a European one learns to understand other restaurant guests by learning the rule-based behaviour characteristic of high-class restaurant situations in Japan (special greetings, taking off shoes, sitting in a special way, etc.). Thus, we not only create person models but also situation models, and our understanding of others uses both types of model as input and selects the model most helpful for evaluating the other person's behaviour. In the following, we explicate in two steps, first, why situation models are necessary and, second, how person models and situation models interact in the social evaluation process (figure adapted from Newen 2015a, p. 21).



The relevance of situation models is based on the observation that situation models are sufficient to understand others, if we need not account for the individual person we aim to understand. If one is not interested in another as an individual, but merely as another agent in a situation, the situation itself often offers sufficient information to predict the behaviour of the people and to coordinate one's own action with theirs, e.g. many shopping interactions are of this type. Furthermore, situation models allow us to predict the behaviour of all people fulfilling typical roles in the situation, e.g. the role of the guest, the seller or the cleaning person within a restaurant. Since humans frequently need to coordinate actions with many persons in the same situation, understanding others on the basis of situation models is a very important tool for life in larger groups. The distinction between person models and situation models is also captured in the difference between reason explanations which focus on consideration of an individual, on the one hand, and causal history explanations which highlight the relevant situation and how it developed, on the other hand (Malle et al. 2006, Fiebich 2015).

Thus, a full-fledged theory of understanding others needs to include situation models as well as basic ideas of the interdependence of personal models and situation models. A situation model is a pattern constituted by a sequence of typical activities or events in a type of situation involving human agents whereby the agents are only represented in an unspecific way comparable to variables in logic. Paradigmatically, this includes situations, such as entering a fast-food restaurant to arrange your lunch, entering a class room to participate in a university seminar, entering a bar to meet friends, etc. We need situation models to coordinate quickly with others according to social expectations in such situations, even if we do not know the persons involved. Of course, we need to account for the fact that situation models

vary intensely across cultures, e.g. how you have to behave in a restaurant is quite different in Japan and the United States.

Finally, we will briefly illustrate the second aspect, i.e. how person models and situation models interact. If we consider once more the restaurant situation, then a typical case could be that while waiting in the queue to order lunch, one starts to communicate with the person in front. This immediately initiates a basic person perception (Macrae & Quadflieg 2010; Newen 2015a) which leads to the creation of a person model, at least, in the working memory. Whether it is transferred into long-term memory depends on attentional features that rely on the estimated relevance of this interaction for future life. For instance, if the person in front impresses us a lot or if we discover that she will start working in the same company, this would lead to the creation of an explicit person model, i.e. a person image, which is enriched step by step during each encounter. If we have a minimally rich person model, it is cognitively economical to rely on this person model whenever interacting with the person. It allows for much better predictions because we can account not only for the general information concerning a situation type, but also for the more fine-grained information about this very person.

Situation models can be enriched by person models and the other way around: a situation model is a pattern constituted by a sequence of typical activities or events in a type of situation involving human agents whereby the agents are only represented in an unspecific way comparable to variables in logic. Person models can enrich these variables in a way that can be compared to substituting a variable by a logical constant (by fitting the person representations into the unspecific agent slots of the event structure). If one engages in a social understanding with an activation of a person model, it can of course be naturally enriched by integrating the person model into a relevant situation model (by including person representations into an event sequence). The richest understanding of a social situation is possible, if we have a detailed situation model. For instance, when we are at a formal birthday party, and we know each member of the party, we can enrich the situation model with all relevant person models. This allows us to deliver detailed explanations and make detailed predictions. To sum up: We rely on both person models and situation models. Situations models are more important for basic social perceptions of situation including the agents involved in the situation (but ignoring them as individuals), while person models are especially fruitful, if an understanding of the individuals themselves is relevant. The richest understanding demands a combination of both and their integration in a structurally fitting manner.

3.3 General Profile and Advantages of PMT

The person model theory (PMT) contains both, a theory about the epistemic strategies involved in social understanding and a theory about the organizational structure of relevant background information, either in

person models or situation models. The resulting general picture of social understanding is the following. Social understanding needs the activation of at least one epistemic strategy. Epistemic strategies can be used individually or in combination as they gradually occur in ontogeny. Epistemic strategies are activated in a particular situation by social cues. This activation can be based on a single social cue, e.g. biological motion, but typically, it relies on the perception of many social cues at once, e.g. in the case of emotion recognition based on facial expression, body posture, gestures. Moreover, in the activation of a particular epistemic strategy typically person models and/or situation models are involved. Social understanding does not merely rely on directly perceivable social cues but also on background information organized in models. Person models and situation models of certain entities also unfold ontogenetically (Newen & Marchi 2016) and are systematically enriched by the new information a person receives. Thus, these models are not rigid but dynamically developing.⁸ This enables flexible and reliable social registration, prediction and evaluation. The interaction of epistemic strategies, person models and situation models enables thereby a great variety of types of social understanding. In a simplified overview, we distinguished ontogenetically three main types of social understanding:

- (i) *online understanding of others* realized as coordinated interaction or as participating in joint attention and joint action, typically based on intuitive epistemic strategies like direct perception or low-level simulation in combination with person model information, e.g. a child smoothly interacting with its mother activating an interaction schema or joint attention concerning an object.
- (ii) *understanding of others as rule-followers* realized as expecting others to follow rules which are constitutive for members of a social group in a specific situation. This is based on person models for types of individuals which are also called person models of groups (Newen 2015a), e.g. when we recognize a person as a member of a soccer club, we expect him to be a good soccer player preparing for a game when it is time to do so or to help preparing a club party when the celebration of the local championship is announced. These examples illustrate that person models of groups (types of individuals) are essentially involved and they have to be combined with situation models to activate specific expectations. Furthermore, some epistemic strategy has to be activated and this can be any of the available strategies. Such a flexible use of the most efficient epistemic strategy is also presupposed for the third type of social understanding:

8 This implies that the epistemic strategies are not only used to register and evaluate a social situation but they are at the same time a tool of adjusting the contents of person models and situation models.

- (iii) *understanding of others as having an individual mindset of attitudes realized by the attribution of beliefs, desires, hopes, fears, etc. to others in a certain situation. This is based on person models of individuals in a specific situation. Thus, persons are represented as having individual beliefs and desires such that e.g. they can be predicted not to behave according to a relevant social roles they have in a situation but according to their individual mindset which can differ from relevant social expectations.*

In the scope of this article, the PMT cannot be outlined in more detail. However, it has been illustrated that PMT clearly differs from its competitors. PMT can account for the plurality and development of epistemic strategies employed by a person in different situations with regard to different agents and in different stages of ontogenetic development. In addition, the introduction of the person model enables the understanding of several important aspects which at least one of the competitors fails to account for:

1. The person model theory can convincingly account for the difference between understanding a complete stranger by relying on a situation model and understanding a well-known familiar person by relying on a rich and more specific person model. No other theory can account for the systematic understanding of individual idiosyncrasies and the relevance of cultural schemata of how to behaviour in a particular context. On the contrary, person models and situation models can do the job.
2. By appealing to the distinction between implicit and explicit person models, PMT can account for the difference between basic or intuitive understanding and complex or theory-based understanding of others which is underdeveloped in TT.
3. With the difference between a person model of oneself and person models of others, PMT can account for an understanding of others which goes beyond the own-self model as the sole source of understanding others, contrary to ST. PMT can account both for an understanding of others based on the self-model and an understanding of others based on the person model of other individuals or types of individuals which can be radically different from the self-model.
4. PMT differs from IT, since it addresses not only basic online understanding, but also offline social understanding.
5. Furthermore, with the outlined dynamics between situation models and person models, PMT especially offers a tool to account for situational and personal features as well as the cultural variation of their relevance. Thus, it seems correct to call PMT a new approach, not just a variant of an existing one.

3.4 Challenges

PMT is a new framework to account for understanding others which confers several advantages compared to the alternative accounts. Nevertheless, it is still accompanied by open questions and challenges. This includes among others a more precise investigation of how the different epistemic strategies interact and under which conditions one strategy is preferred in case it conflicts with another. Moreover, there is a need to clarify further how person models are individuated and how they are cognitively implemented. While in the published work, there is a description of an fMRI study of Hassabis et al. (2013) that provides evidence for person models as models of personality types (person schemata for groups), it would be helpful to provide similar evidence concerning the implementation of person models of individuals. PMT also needs an explication of the borderline between person models of groups which are already general and rather general knowledge of folk psychological rules as described by TT. Finally, a detailed description of the interaction between situation models and person models is needed as they strongly influence each other: a person often has different dispositions to behave, to the point of virtually different personality traits showing up depending on the situation. In a job situation a person may be extremely harsh, while being friendly in a family context.

4. Concluding remarks

We discussed the four main theories of understanding others: ST, TT, IT and PMT. While the first three accounts – which have indeed been under discussion in the philosophical literature for quite some time – reveal critical gaps, PMT offers a promising attempt to close these gaps, albeit still having open questions that its defenders have to answer. The future of the debate about social cognition will tell which theory is the most fruitful framework and how it should be unfolded to deliver the most adequate descriptions and predictions. This is clearly an interdisciplinary challenge which requires the combination of insights at least from philosophy, psychology, psychiatry and neuroscience. The current state of knowledge indicates the need for a multidimensional and flexible understanding of human mindreading which involves individuals, groups, cultures and situations.

References

- Adolphs, R., D. Tranel, H. Damasio & A. Damasio 1994. “Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala.” *Nature* 372 (6507): 669–72.
- Andrews, K. 2012. *Do apes read minds? Towards a new folk psychology*. Cambridge (MA): MIT Press.

- Apperly, I. & S. Butterfill 2009. "Do humans have two systems to track beliefs and belief-like states?" *Psychological Review* 116 (4): 953–970.
- Baron-Cohen, S. 1995. *Mindblindness – An Essay on Autism and Theory of Mind*. Cambridge (MA); London: The MIT Press.
- Becchio, C., M. Adenzato & B. G. Bara 2006. "How the brain understands intention: different neural circuits identify the componential features of motor and prior intentions." *Consciousness and Cognition* 15 (1): 64–74.
- Bertin, E. & T. Striano 2006. "The still-face response in newborn, 1.5-, and 3-month-old infants." *Infant Behavior and Development* 29 (2): 294–297.
- Birch, S.A. & P. Bloom 2003. "Children are Cursed: An Asymmetric Bias in Mental-State Attribution." *Psychological Science* 14 (3): 283–6.
- Bohl, V. 2015. "Continuing debates on direct social perception: Some notes on Gallagher's analysis of 'the new hybrids'." *Consciousness and Cognition* 36: 466–71.
- Butterfill, S. A. 2009. "Seeing causings and hearing gestures." *Philosophical Quarterly* 59 (236): 405–28.
- Camerer, C., G. Loewenstein & M. Weber 1989. "The Curse of Knowledge in Economic Settings: An Experimental Analysis." *Journal of Political Economy* 97 (5): 1232–54.
- Carruthers, P. 2015. "Perceiving Mental States." *Consciousness and Cognition* 36: 498–507.
- Davies, M. & T. Stone 1995. *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell Publishers.
- De Bruin, L. & A. Newen 2012. "An Association Account of False Belief Understanding." *Cognition* 123 (2): 240–259.
- De Bruin, L., M. van Elk & A. Newen 2012. "Reconceptualizing Second-Person Interaction." *Frontiers in Neuroscience* 6: 151.
- De Jaegher, H. and Di Paolo, E. 2007. "Participatory sense-making: An enactive approach to social cognition." *Phenomenology and the Cognitive Sciences* 6 (4): 485–507.
- Di Paolo, E. & H. De Jaegher 2012. "The interactive brain hypothesis." *Frontiers in Human Neuroscience* 6: 1–16.
- Di Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese & G. Rizzolatti 1992. "Understanding motor events: a neurophysiological study." *Experimental Brain Research* 91 (1): 176–80. Retrieved from <http://link.springer.com/article/10.1007/BF00230027>
- Ekman, P. & W. V. Friesen 1971. "Constants across culture in the face and emotion." *Journal of Personality and Social Psychology* 17: 124–129.

- Fiebich, A. 2015. *Varieties of Social Understanding*. Münster: mentis Verlag.
- Fiebich, A. & Coltheart, M. 2015. "Various ways to understand other minds. Towards a pluralistic approach to the explanation of social understanding." *Mind and Language* 30 (3): 235–58.
- Fisher, J. C. 2006. "Does Simulation Theory Really Involve Simulation?" *Philosophical Psychology* 19 (4): 417–32.
- Frith, U., & Frith, C. D. 2003. "Development and neurophysiology of mentalizing." *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 358 (1431): 459–473.
- Furley, P., T. Moll & D. Memmert 2015. "'Put your Hands up in the Air'? The interpersonal effects of pride and shame expressions on opponents and teammates." *Frontiers in psychology* 6: 1361.
- Gallagher, S. 2001. "The practice of mind: Theory, simulation, or interaction?" *Journal of Consciousness Studies* 8 (5–7): 83–107.
- Gallagher, S. 2007. "Simulation Trouble." *Social Neuroscience* 2 (3&4): 653–65.
- Gallagher, S. 2008. "Direct perception in the intersubjective context." *Consciousness and Cognition* 17 (2): 535–43.
- Gallese, V. & A. Goldman 1998. "Mirror neurons and the simulation theory of mind-reading." *Trends in Cognitive Sciences* 2 (12): 493–501. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1364661398012625>
- Gao, T., Newman, G. E., and Scholl, B. J. 2009. "The psychophysics of chasing: A case study in the perception of animacy." *Cognitive Psychology* 59 (2): 154–79.
- Godfrey-Smith, P. 2005. "Folk psychology as a model." *Philosophers' Imprint*, 5 (6): 1–16.
- Goldman, A. I. 1989. "Interpretation Psychologized." *Mind & Language* 4 (3): 161–85.
- Goldman, A. I. 1992. "In Defense of the Simulation Theory." *Mind & Language* 7 (1&2): 104–19.
- Goldman, A. I. 2006. *Simulating Minds – The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.
- Goldman, A. I. & L. C. Jordan 2013. „Mindreading by simulation: The roles of imagination and mirroring." In *Understanding Other Minds*, 3rd edition, edited by S. Baron-Cohen, M. Lombardo & H. Tager-Flusberg, 448–66. Oxford, Oxford University Press.
- Goldman, A. I. & N. Sebanz 2005. "Simulation, mirroring, and a different argument from error." *Trends in Cognitive Sciences* 9 (7): 320; author reply 321.

- Gopnik, A. & H. M. Wellman 1992. "Why the Child's Theory of Mind Really Is a Theory." *Mind & Language* 7 (1&2): 145–71.
- Gordon, R. M. 1986. "Folk Psychology as Simulation." *Mind & Language* 1 (2): 158–71.
- Gordon, R. M. 1992. "The Simulation Theory: Objections and Misconceptions." *Mind & Language* 7 (1&2): 11–34.
- Hassabis, D., Spreng, R. N., Rusu, A. A., Robbins, C. A., Mar, R. A., and Schacter, D. L. 2013. "Imagine all the people: how the brain creates and uses personality models to predict behavior." *Cerebral Cortex*, March 5, 2013. doi:10.1093/cercor/bht042.
- Herschbach, M. 2012. "On the role of social action in social cognition: a mechanistic alternative to enactivism." *Phenomenology and the Cognitive Sciences* 11 (4): 467–486.
- Hutto, D. D. 2007. "The narrative practice hypothesis: origins and applications of folk psychology." *Royal Institute of Philosophy Supplement* 60: 43–68.
- Hutto, D. 2008. *Folk-psychological narratives*. Cambridge (MA): MIT Press.
- Hutto, D. & S. Gallagher 2008. "Understanding others through primary interaction and narrative practice." In *The Shared Mind: Perspectives on Intersubjectivity*, Converging Evidence in Language and Communication Research 12, edited by J. Zlatey, T. Racine, C. Sinha & E. Itkonen, 17–38. Amsterdam: John Benjamins Publishing Company.
- Keysers, C. & V. Gazzola 2009. "Unifying Social Cognition." *Progress in Brain Research* 156: 379–401.
- Kilner, J. M., K. J. Friston & C. D. Frith 2007. "The mirror-neuron system: a Bayesian perspective." *Neuroreport* 18 (6): 619–23.
- Kuzmanovic, B., Schlibach, L., Lehnhardt, F., and Vogeley, K. 2011. "A matter of words: Impact of verbal and nonverbal information on impression formation in high-functioning autism." *Research in Autism Spectrum Disorders* 5 (1): 604–13.
- Macrae, C. N. & Bodenhausen, G. V. 2000. Social cognition: Thinking categorically about others. *Annual Review of Psychology*, 51, 93–120.
- Macrae, C. N. & Quadflieg, S. 2010. Perceiving people. In S. Fiske, D. T. Gilbert & G. Lindzey (Eds.) *Handbook of Social Psychology*. New York, NY: McGraw-Hill, 428–463.
- Malle, B. F. 2006. "The actor-observer asymmetry in attribution: a (surprising) meta-analysis." *Psychological Bulletin*, 132 (6), 895–919.
- Marchi, F. & A. Newen 2015. "Cognitive penetrability and emotion recognition in human facial expressions." *Frontiers in Psychology* 6 (828). doi: 10.3389/fpsyg.2015.00828

- 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.
- Maibom, H. L. 2009. "In defence of (model) theory theory." *Journal of Consciousness Studies* 16 (6–8): 360–78.
- Newen, A. 2014. "Selbst- und Fremdverstehen: die Personenmodelltheorie als Analyserahmen für mentale Störungen." In *Verleumdung und Verrat: Dissoziative Störungen bei schwer traumatisierten Menschen in Folge von Vertrauensbrüchen*, edited by Vogt, R., Kröning: Asanger, 209–218.
- Newen, A. 2015a. "Understanding Others. The Person Model Theory" In *Open MIND*, 26, edited by J. M. Metzinger & T. Windt, 1–28. Frankfurt am Main: MIND Group. [Open.mind.net/](http://open.mind.net/) doi: 10.15502/9783958570320, 1–28 [Printed version: Newen, A. (2016): "Understanding Others – The Person Model Theory." In: Metzinger, T. & Windt, J.M. (eds.): *Open MIND. Philosophy and the Mind Sciences in the 21st Century*, Cambridge MIT Press, S. 1049–1076].
- Newen, A. 2015b. "A Multiplicity View for Social Cognition: Defending a Coherent Framework." In: *Open MIND*, 26, edited by J. M. Metzinger & T. Windt. Frankfurt am Main: MIND Group. <http://open-mind.net/> doi: 10.15502/9783958570320 [Printed Version: Newen, A. (2016): A Multiplicity View for Social Cognition: Defending a Coherent Framework – A reply to Lisa Quandt. In: Metzinger, T. & Windt, J.M. (Hrsg.): *Open MIND. Philosophy and the Mind Sciences in the 21st Century*, Cambridge MIT Press, 1095–1102].
- Newen, A. 2017. "Defending the liberal-content view of perceptual experience: Direct social perception of emotions and person impressions." *Synthese* 194 (3): 761–785. doi: 10.1007/s11229-016-1030-3.
- Newen, A. 2018. "The Person Model Theory and The Question of Situatedness of Social Understanding. In: Newen, A., de Bruin, L., Gallagher, S. (eds): *Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press, 469–492.
- Newen, A. & F. Marchi 2016. "Concepts and their organizational structure." In *Concepts and Categorization*, edited by D. Hommen, C. Kann and T. Ossenbald, Münster: mentis Verlag, 197–227.
- Newen, A. & T. Schlicht 2009. "Understanding Other Minds: A criticism of Goldman's simulation theory and an outline of the Person Model Theory." *Grazer Philosophische Studien* 79: 209–42.
- Newen, A. & P. Vetter 2016. „Why cognitive penetration of our perceptual experience is still the most plausible account." *Consciousness and Cognition* 47: 26–37.

- Newen, A. and Vogeley, K. 2011. "Den anderen verstehen." *Spektrum der Wissenschaft* 8/2011.
- Newen, A., A. Welpinghus & G. Juckel 2015. "Emotion Recognition as Pattern Recognition: The Relevance of Perception." *Mind & Language* 30 (2): 187–208.
- Neufeld, E., Brown, E. C., Lee-Grimm, S.-I., Newen, A., Brüne, M. (2016). Intentional Processing Results from Automatic Bottom-Up Attention: An EEG-Investigation into the Social Relevance Hypothesis Using Hypnosis. *Consciousness and Cognition* 42, 101–112. doi: 10.1016/j.concog.2016.03.002
- Nickerson, R. S. 1999. "How we know – and sometimes misjudge – What others know: Imputing one's own knowledge to others." *Psychological Bulletin* 125 (6): 737–59.
- Onishi, K. H. & R. Baillargeon 2005. "Do 15-month-old infants understand false beliefs?" *Science* 308 (8), 255–258.
- Pacherie, E. 2005. "Perceiving intentions." In *A explicação da interpretação humana*, edited by J. Sàágua, Lisbon: Edições Colibri, pp.401–14.
- Perry, J. 1990. "Self-Notions." *Logos* 11: 17–31.
- Rakoczy, H., Warneken, F., Tomasello, M. 2008. "The Sources of Normativity: Young children's awareness of the normative structure of games." *Developmental Psychology* 44 (3): 875–881.
- Recanati, F. 2012. *Mental Files*. Oxford, UK: Oxford University Press.
- Rizzolatti, G. & L. Craighero 2004. "The mirror-neuron system." *Annual Review of Neurosc.* 27: 169–92.
- Rutherford, M. D. and Kuhlmeier, V. A. 2013. *Social perception: Detection and interpretation of animacy, agency, and intention*. Cambridge (MA): MIT Press.
- Saxe, R. 2005. "Against simulation: the argument from error." *Trends in Cognitive Sciences* 9 (4): 174–9.
- Scheler, M. 1954. *The Nature of Sympathy*. New Haven: Yale University Press.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., Vogeley, Kai. 2013. "Toward a second-person neuroscience." *Behavioral and Brain Sciences* 36: 393–462. doi:10.1017/S0140525X12000660
- Siegel, S. 2009. "The visual experience of causation". *Philosophical Quarterly* 59 (236): 519–40.
- Sinigaglia, C., G. Rizzolatti 2008. *Mirrors in the Brain: How Our Mind Share Actions and Emotions*. Oxford: Oxford University Press.
- Spaulding, S. 2010. "Simulation theory." *Wiley Interdisciplinary Reviews: Cognitive Science* 1 (4): 527–38.

- Sodian, B., C. Thoermer & U. Metz 2007. "Now I see it but you don't: 14-month-olds can represent another person's visual perspective." *Developmental Science* 10 (2): 199–204.
- Sorce, J., R. Emde, J. Campos & M. Klinnert 1985. "Maternal Emotional Signaling: Its Effect on the Visual Cliff Behavior of 1-Year-Olds." *Developmental Psychology* 21 (1): 195–200.
- Stich, S. & S. Nichols 1995. "Second Thoughts on Simulation." In *Mental Simulation: Evaluations and Applications*, edited by A. Stone & M. Davies, 87–108. Oxford: Blackwell.
- Tomasello, M. 1995. "Joint attention as social cognition." In *Joint attention: Its origins and role in development*, edited by C. Moore & P. J. Dunham, 103–130. Hillsdale: Lawrence Erlbaum.
- Tomasello, M. 1999. *The cultural origins of human cognition*. Cambridge (MA): Harvard University Press.
- Trevarthen, C. B. 1979. "Communication and cooperation in early infancy: A description of primary intersubjectivity." In *Before speech: The Beginning of Interpersonal Communication*, edited by M. Bullowa, 321–48. Cambridge: Cambridge University Press.
- Trevarthen, C. B. 2006. "The concepts and foundations of intersubjectivity." In *Intersubjective Communication and Emotion in Early Ontogeny*, edited by S. Braten, 15–46. Cambridge: Cambridge University Press.
- Trevarthen, C. B. & P. Hubley 1978. "Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year." In *Action, gesture and symbol: The emergence of language*, edited by A. Lock, 183–229. London: Academic.
- Tversky, A. & D. Kahneman 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.
- Tye, M. 1995. *Ten problems of consciousness. A representational theory of the phenomenal mind*. Cambridge, MA: MIT Press.
- Vetter, P. & Newen, A. 2014. "Varieties of cognitive penetration in visual perception." *Consciousness & Cognition* 27: 62–75. Doi:10.1111/j.1468-0017.2006.00298.x.
- Vogeley, K. 2012. *Anders sein – Hochfunktionaler Autismus im Erwachsenenalter*. Weinheim: Beltz-Verlag.
- Vogeley, K., P. Bussfeld, A. Newen, S. Herrmann, F. Happé, P. Falkai, W. Maier, N. J. Shah, G. R. Fink & K. Zilles 2001. "Mind reading: neural mechanisms of theory of mind and self-perspective." *Neuroimage* 14 (1 PT 1): 170–81.
- Vogeley, K. & A. Newen 2002. "Mirror neurons and the self construct." In *Mirror Neurons and the Evolution of Brain and Language*, edited by Maxim

- I. Stavenov and Vittorio Gallese, 135–150. Amsterdam: John Benjamins Publishing.
- Vogeley, K., L. Schilbach & A. Newen 2013. “Soziale Kognition.” *Interdisziplinäre Anthropologie* 1: 13–40.
- Weinberg, M. K., M. Beeghly, K. L. Olson & E. Tronick 2008. “A Still-face Paradigm for Young Children: 2½ Year-olds’ Reactions to Maternal Unavailability during the Still-face.” *Journal of Developmental Processes* 3 (1): 4–22.
- Wicker, B., C. Keyers, J. Plailly, J. P. Royet, V. Gallese & G. Rizzolatti 2003. “Both of us disgusted in *My insula*: the common neural basis of seeing and feeling disgust.” *Neuron* 40 (3): 655–64.
- Wimmer, H. & J. Perner 1983. “Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception.” *Cognition* 13: 103–28.
- Zahavi, D. 2011. “Empathy and direct social perception: A phenomenological proposal.” *Review of Philosophy and Psychology* 2 (3): 541–58.