Multiplicity Needs Coherence – Towards a Unifying Framework for Social Understanding

A Commentary on Albert Newen

Lisa Quadt

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

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

Commentator

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

Target Author

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

Editors

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

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

1 Introduction

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

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

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

2 The multiplicity view

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

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

Newen claims that

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

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

3 Multiplicity needs coherence

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5 1-3sE– Levels of social embodiment

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

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

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

5.1 Third-order social embodiment (3sE)

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

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

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

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

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

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

5.2 Second-order social embodiment (2sE)

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

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

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

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

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

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

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

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

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

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

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

5.3 First-order social embodiment (1sE)

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

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

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

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

6 Conclusion

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

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

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

Acknowledgements

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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