

---

# Action-Oriented Predictive Processing and Social Cognition

Lisa Quadt

---

The research field on social cognition currently finds itself confronted with two conflicting theoretical camps, cognitivism and enactivism. In their most extreme formulations, the former claims that mindreading skills exhaust our social cognitive capacities, while the latter stresses the sufficiency of interaction and embodiment. My aim is to find a middle position that provides the basis for discussing social cognition as interactive and embodied, while remaining in non-radical territory.

This can be achieved by situating social cognition within the framework of action-oriented predictive processing (Clark 2013). Specifically, I propose three conceptual tools, namely (1) embodied social inference (EmSI), (2) action-oriented predictions (a-o predictions) (Clark 2016), and (3) interactive inference (InI).

The first concept of EmSI refers to the more general term of “embodied inference” (Friston 2012), which means that an organism’s morphology incorporates the demands of its environment. This idea can be applied to the social realm, in the sense that the kind of body an individual has constrains the kind of social interaction they can engage in. While humans, for example, can exploit their speech apparatus for communication, ants instead rely on their pheromone system. The body of an individual thus also constrains social cognitive skills and can be said to play a crucial role for interactions. This becomes obvious when considering the second concept of “action-oriented predictions”. The basic idea is that the job of a predictive model is to distribute the cognitive workload and recruit embodied action whenever possible. Here too the body plays an indispensable role in that it realizes prediction error minimization by engaging with the external world via active inference. Related to this idea is the last concept of “interactive inference”. I claim that interaction plays the same role for social cognition as action does for general cognition — namely gathering information about the *social environment* and thus actively sculpting not only one’s external, but also internal environment. InI can be described as the minimization of prediction error while navigating the social environment. It serves to actively sample proof for predictions or to disambiguate competing models about the other.

In what I call replicative interactive inference (RIInI), the bodily state (e.g., posture, movements) of another person is mimicked in order to supplement exteroceptive information about them with interoceptive and proprioceptive information. Mimicry, synchronization and automatic imitation are instances of RIInI that function to make predictions about the other more precise by increasing the number of signal sources that yield relevant information.

Secondly, complementary interactive inference (CIInI) refers to changing one’s internal or external environment in response to the other person. It serves to either regulate the other’s current state (e.g., mothers lowering their body temperature to cool down their infant’s feverish body; Nyqvist et al. 2010), or to evoke further behavioral responses that then serve as additional exteroceptive input (e.g., using gestures to express one’s uncertainty).

These conceptual tools can serve to alleviate the tension between enactivist and cognitivist theories. The present proposal thereby enables a dialogue about social cognition as an interactive and embodied process.

## Keywords

Active inference | Action-oriented predictive processing | Action-understanding | Embodied inference | Interaction | Social cognition

## 1 Introduction

The topic I will pursue in the current paper concerns the implications that predictive processing (PP; Hohwy 2013) has for research on social cognition. More specifically, I will discuss the possibilities that action-oriented PP (Clark 2013) holds for beginning to build a comprehensive theoretical framework for social cognition.

The paper is divided into two parts. In the first part I argue that a fresh view on the phenomenon is needed because the research field of social cognition currently finds itself confronted with two conflicting theories, viz., cognitivism and phenomenology/enactivism (henceforth *phenactivism*<sup>1</sup>). In their radical<sup>2</sup> formulations, the former claims that so-called mindreading skills — i.e., simulation (Gallese and Goldman 1998) and theoretical inference (Gopnik and Wellman 1992) — exhaust our social cognitive skills. The latter, on the other hand, emphasizes that social cognition entails embodied interaction and even claims that interaction patterns may constitute social cognitive processes (De Jaegher and Di Paolo 2007). This situation is problematic because both theoretical camps have their problems, which leave them unfit for serving as the basis for a comprehensive theory of social cognition. Phenactivism, it will be claimed, has neither a sound conceptual nor empirical basis, and therefore is unable to provide the means for a theoretical framework in which social cognition can be embedded. Cognitivism, on the other hand, neglects the issues of interaction and embodiment almost entirely and thus draws an incomplete picture of the manifold phenomenon of social cognition. At the same time, however, both theories make valuable positive proposals that should be considered in a theory of social cognition.

In the second part of the paper I argue that what is required to alleviate this tension is a new view on social cognition that integrates insights from both sides of the theoretical spectrum while remaining in non-radical territory. Action-oriented PP, as will be described in section 3, provides the conceptual tools to do just that. The term has been introduced by Clark (Clark 2013) to capture the idea that PP unifies action, perception and cognition in one theoretical framework. Perception and action are thought to follow the same computational principles and to crucially depend on each other in their joint mission to minimize prediction error. Where perception generates prior expectations about the unfolding of sensory consequences, action functions to fulfill these expectations by sampling the world (cf. Friston 2009, p.12). In this scheme, perception cannot do without action, and *vice versa*. Three aspects of action-oriented PP will be discussed in order to later embed them into the context of social cognition, viz., embodied inference (Friston 2010), action-oriented predictions (Clark 2016), and active inference (Clark 2015b).

In section 4, I aim to exploit this picture of the mind as drawing both on internal modeling and engagement of the environment by embodied agents in order to make it fruitful for research on social cognition. I will introduce three conceptual tools that shall provide a conceptual basis for further research. First, the notion of embodied social inference (EmSI) is presented. EmSI is meant to capture the idea that the very physiology of an agent constrains their range of social interactions. Secondly, the concept of action-oriented predictions is applied to and made fruitful for social cognition. Lastly, I introduce the term interactive inference (InI) in order to be able to assign an important role to interactive processes for social cognition.

- 1 Enactivism puts much emphasis not only on the body, but especially on interaction as a potentially constitutive element of social cognition. 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 to build a proper account of cognition (Varela et al. 1993), phenomenology is mostly seen as a description of experiential phenomena (Gallagher 2008). I use the word ‘phenactivism’ to describe views that merge phenomenology and enactivism. Since they share fundamental premises (Quadt 2015) they can be subsumed under this concept.
- 2 I use the term radical in the sense of ‘extreme’, not in the sense of ‘anti-representationalist’. It is important to notice that I will here describe only one and a rather radical version of each theoretical strand. Of course, either theory has been presented in various ways and with differing assumptions, some more and some less radical. Presenting the multitude of versions of each theory is neither necessary nor within the scope of this paper.

## 2 Phenactivism vs. Cognitivism

### 2.1 Conceptual Clarification

The goal of this section is to lay out the basic claims of phenactivism and cognitivism and to discuss their different approaches to general and social cognition. It is claimed that the two accounts can be seen as marking the endpoints of a theoretical spectrum, both of them providing valuable insights and assumptions about the nature of the human mind.

I will start with clarifying the concepts of “phenactivism” and “cognitivism”. Both terms refer to specific accounts of cognition that hold a distinct set of metaphysical, methodological, and epistemological background assumptions. Most readers will be more familiar with cognitivist views on the mind, since these have not only been the prevalent accounts since the rise of cognitive science but still continue to form the theoretical background of most researchers in the field. My description of cognitivism will thus be rather short and my focus will be on disentangling the more obscure and less famous notion of what I call phenactivism. Useful definitions of each term are provided by De Bruin and Kästner ([de Bruin et al. 2012](#), p. 542-543) and serve to give a first idea of what they amount to:

Classic Cognitivism (COG): The mind is basically an intracranial information processing system manipulating (sub-)symbolic representations; cognition essentially is this computational process.

Enactive Cognition (ENAC): Rather than a representational process, cognition is a process of sense-making that emerges from the dynamic online interaction or ‘coupling’ between autonomous agents and the environment in which they are embedded.

In other words, while cognitivism describes the human mind as a computational device that can be found exclusively inside the skull of an individual and operates on representations, (ph)enactivism claims quite the opposite; the mind is neither inside nor outside the individual but instead emerges within the *relation* of agent and environment. In the following, I will unpack each term further.

### 2.2 Cognitivism

Classic or radical cognitivism has been described above as viewing the mind as an entirely internal device that operates on representations in a specific, symbolic format. The notion of representations is as central as the claim that cognition is skull-bound and computational. This theory is the metaphysical and methodological background for so-called mindreading theories of social cognition. These theories are traditionally simulation-theory and theory-theory and in principle describe internal processes of a certain kind that underlie the inference of mental states of others. In their most extreme formulations, these theories state that social cognitive skills are mindreading skills and thereby draw a fundamentally individualistic, internalist, and representationalist picture of the phenomenon.

The role of social interaction and embodiment in radical cognitivist views is quickly explained. Mindreading theories have paid little attention to social interaction and embodiment and how these factors could influence, change, or even constitute social cognition. However, it should be noted that they are not obliged to deny that both are important factors for social understanding ([Overgaard and Michael 2013](#)). This is especially obvious from the experimental paradigms that are used to investigate mindreading skills. Typically, social stimuli consist of the picture of another person or a video of this person executing a specific action (e.g., [Iacoboni et al. 2005](#); [Wicker et al. 2003](#)). While this kind of experimental design is well controlled, it lacks ecological validity since it situates participants in rather unrealistic situations.

Taken together, radical cognitivist theories foster a rather inflexible view on the mind as an input-output device. This view is then transferred to the social realm, drawing a picture of social cognition that ignores the fact that social encounters involve embodied agents that engage in interactions with each other.

## 2.3 Phenactivism

The shortcomings of cognitivism discussed above motivated phenactivists to find an alternative perspective that considers not just the brain in the skull but also the organism in the environment. Phenactivism describes the mind as *relational*, as emerging in the interaction of agent and environment. In the early 1990s, Varela, Thompson and Rosch (Varela et al. 1993) published their book *The Embodied Mind* in which they aimed to provide a non-cognitivist, alternative model of the mind. Their motivation was to criticize the view that describes mental processes as computations and the manipulation of representations. Such a model is said to be unsatisfactory, since it lacks a pragmatic approach to cognition and fails to integrate an inherent connection between mind and life (cf. Thompson 2010, p. 12).

A radical cognitivist picture of the mind depicts mental processing as fully internal and will thus not attribute any decisive role to the body. Phenactivists, however, adopt a rejection of the distinction between inner and outer, claiming that the first mistake to make in thinking of cognition is to assume that it has a location which is found either inside or outside the skull (Arnau et al. 2014). This point is most crucial for understanding the difference between phenactivism and cognitivism. Even tamer versions of cognitivism, which state that the mind can be extended to brain-external structures, are still fundamentally different from phenactive views.

While cognitivism places epistemic mechanisms within the skull and attributes a mere input-role to the external world and an output-role to action, phenactivism ties in both of these elements into the epistemic process. This is captured by the central notion of *sense-making*; within their embodied activity, agents not only actively regulate their coupling with the environment, they thereby establish a perspective onto the world (cf. De Jaegher and Di Paolo 2007 p. 488). Agents thus *create* meaning, there is no passive reception of information which is processed into or in virtue of internal representations which then (potentially) bear meaningful content.

The centrality of interaction is the core assumption of phenactive accounts and builds the starting point for further claims. Social interactions are seen as providing enabling conditions and forming constitutive elements for both the development and maintenance of social skills (De Jaegher and Di Paolo 2007; De Jaegher et al. 2010; Di Paolo and De Jaegher 2012). In order to expound this view, the claim is couched in theoretical terms of general phenactivism. Empirical set-ups, such as the perceptual crossing paradigm (Auvray et al. 2009) are assumed to corroborate these theoretical aims.

At this point it will be helpful to look at how proponents of the theory conceive of interaction. Here is a definition that is now generally accepted:

Social interaction is the regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter's scope can be augmented or reduced). (De Jaegher and Di Paolo 2007, p. 493)

In other words, interactions are viewed as building autonomous systems which then are irreducible to local mechanisms physically realized within the individuals involved. Two systems are furthermore said to be coupled when their behavior and mental states depend on each other.

The concept of 'participatory sense-making' was introduced to capture these ideas. De Jaegher and colleagues (De Jaegher and Di Paolo 2007, p. 497) define the term as "the coordination of intentional activity in interaction, whereby individual sense-making processes are affected and new domains of social sense-making can be generated that were not available to each individual on her own." Together with the definition of interaction given above, this means that individuals 'merge' into one interactive, autonomous system. Since sense-making can be seen as the phenactive term for cognition (Thompson 2010), these claims boil down to the statement that interacting individuals, mutually and in virtue of

the emergent interaction dynamics, constitute (at least part of) their social cognitive processes. Social cognition as participatory sense-making then exhibits a *relational* kind of cognition. It is not to be located in either individual's head, brain or even body, but *in between* interacting individuals.

In sum, phenactive views on (social) cognition draw a radically different picture than cognitivist theories and come with radically different premises. This is problematic for several reasons, which I will detail in what follows.

## 2.4 Problems with Phenactivism and Cognitivism

Both phenactivism and cognitivism — in their radical formulations — are ill-suited for providing a comprehensive account on social cognition. The problems that come with a radical cognitivist view on social cognition are rather obvious and mostly refer to the fact that they exclude the importance of interaction and embodiment. While they are good at accounting for high-level phenomena such as explicitly thinking about the causes of another person's behavior, it is mostly ignored that interactions form a context that could change and influence social cognitive processing quite profoundly. If the goal is to find a comprehensive theory of social cognition, a theory that excludes the role of embodied interaction thus is undesirable.

What about the alternative at the other end of the theoretical spectrum? Obviously, phenactivism attributes quite some weight to interaction and embodiment. There is, though, the question of how well their claims are backed up, both conceptually and empirically. In what follows, I will discuss the conceptual and empirical validity of phenactive accounts and conclude that there are many incoherences and uncertainties which leave them unfit to offer a sound theoretical background for social cognition.

To begin with, it appears that phenactivism confuses enabling and constitutive conditions, leaving the phenactivist's claims unclear. A first hint of confusion is found when looking at the taxonomy of possible roles of interaction for a social cognitive process X that De Jaegher and colleagues (De Jaegher et al. 2010, p. 443) have worked out:

Accordingly, given X, and a particular situation in which X occurs: F is a contextual factor if variations in F produce variations in X, C is an enabling condition if the absence of C prevents X from occurring and P is a constitutive element if P is part of the processes that produce X.

As Herschbach (Herschbach 2012) points out, however, it is rather unclear what exactly De Jaegher and colleagues judge to be a constitutive element. For additionally to the characterization given above, they also refer to it as a *part of the phenomenon itself*:

A constitutive element is part of the phenomenon (it must be present in the same time frame as the phenomenon). The set of all the constitutive elements is the phenomenon itself. The presence of these elements is necessary, and therefore also enabling. (De Jaegher et al. 2010, p. 443)

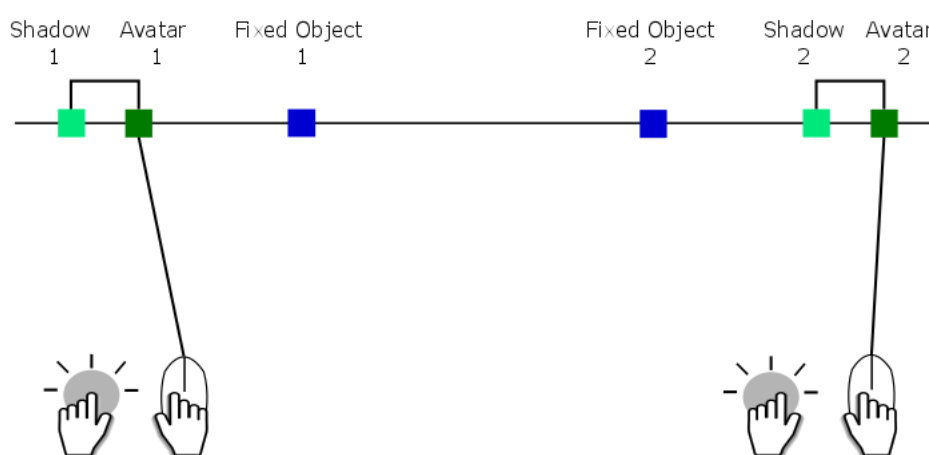
This ambiguity leaves us with two possibilities in which interaction can constitute social cognition: (1) it can either be among those processes that *produce* the phenomenon, but does not have to be a *part of the phenomenon* (e.g., through interacting with her mother, the child learns to 'read' emotions and can later use this skill outside of interactions when she merely thinks about her mother), or (2) interaction constitutes social cognition in the sense that it must be present at the same time as the phenomenon and is a *necessary part* of it (e.g., only when the child interacts with her mother she can 'read' her emotions).

Claim (1) describes a condition that should count as causally enabling, not constitutive. The idea seems to be that interaction *enables* a particular mechanism to *arise* in that it was present as a necessary part of the development of that skill, and therefore it should be called constitutive. This confuses

the concepts profoundly and boils down to the assertion that interaction is an enabling condition and not that it constitutes a phenomenon in the sense that, metaphysically, it is a necessary part of it without which it would not exist. Moreover, the view that being immersed in social interactions — especially from a developmental perspective — enables particular social cognitive skills can in principle be accounted for by any non-phenactive theory that assigns a sufficiently strong role to extra-individual and situational contexts. To see this, consider that human newborns are completely helpless without a caregiver for an extraordinarily long time. Additionally, given some rather anecdotal evidence of children that lacked interactive and emotional engagement in early development and had severe mental as well as bodily impairments (e.g., Zimmer 1989; Bick et al. 2015; Fox et al. 2011), the fact that these contexts play a necessary role for social cognition seems almost trivial. It is questionable whether any theory would reject the assumption that interactive contexts play an enabling role for social cognition.

Further it should be noted that just because something is present in the same time frame as the phenomenon under scrutiny it obviously does not mean that it is part of the phenomenon. However, this is how one could read the quotation above. We can thus draw a first conclusion, stating that phenactive views lack a solid conceptual taxonomy to back up their strong claims in that it is unclear how they identify and separate sets of enabling and constitutive conditions for social cognition. The consequence is that they are left with statements that non-phenactive theories can account for as well.

What is the state of empirical evidence for the claim that interaction constitutes social cognition? Auvray et al's (Auvray et al. 2009) perceptual crossing paradigm is taken as providing an empirical ground for the phenactive position on interaction and as such picks up the idea that there might be something inherent in the interaction dynamics that is irreducible to individual mechanisms. In the experiment, two individuals were blindfolded and had to move their mouse cursor along a line. There were three objects that they could encounter on this line; a fixed object, the avatar of the other person, and the shadow of the other's avatar (Figure 1) Whenever they encountered an object, they would receive tactile feedback. Their task was then to click whenever they thought to have encountered the other's avatar. The results of the study show two things. First, participants were clearly able to distinguish between a fixed and moving object. Secondly, they appeared to favor avatar-avatar encounters, which was obvious from the higher number of these meetings.



**Figure 1:** The perceptual crossing paradigm: In the perceptual crossing paradigm, two participants control an avatar (dark green box) that they can move with a computer mouse along a one-dimensional line with their right hand. The left hand rests on a buzzer, which provides the participants with tactile feedback when their avatar encounters an object in this one-dimensional space. Attached to their avatar is a mobile lure, or “shadow” (light green box), which follows the avatar at a constant, fixed distance. Additionally, there is a fixed, immobile object (blue box) on the line. Participants receive tactile feedback when their avatars encounter the other participant's avatar, their shadow, or the fixed object. The participant's task is to determine when avatars meet; that is, to tell when one participant's avatar encounters the other participant's avatar.

The second finding is said not to be explainable in individual terms and thus to require a non-reductive explanation at the level of collective dynamics. It was found that participants reversed their direction of movement after encountering any object, but only when both avatars meet, both receive a tactile feedback. The result is that, according to the authors, “this co-dependence of the two perceptual activities forms a relatively stable dynamic configuration.” (Auvray and Rohde 2012, p. 3) The fact that an avatar-shadow encounter elicits feedback in only one subject is seen as not allowing the emergence of a stable interaction pattern.

There are, however, ways to interpret the results without referring to interaction dynamics as an emergent macro-structure whose properties substitute a part of individual mechanisms. There are basically three distinct conditions either individual can be in during the task; they can encounter the other person’s avatar, this avatar’s shadow, or the fixed object. Each situation differs with respect to the type of encounter. Importantly, individuals exhibit different behavioral patterns following each encounter. Thus each situation is indeed different, but *in virtue of* the behavior of either individual. It is therefore possible that individuals simply pick up subtle cues in the change of behavior of the other avatar, particularly because the situation in which both participants receive a tactile feedback elicits a different kind of reaction than the other situations.

Further, Froese and colleagues (cf. Froese et al. 2014, p. 8) claim that the results of the paradigm speak for an extendable mind that outsources parts of the cognitive work into the environment. This interpretation is, however, compatible with an extended, yet non-phenactive theory. The same holds for the claim that interaction dynamics influence the cognitive process. The ability to discriminate moving from fixed objects can easily be explained by perceptual learning, the ability to pick up statistical regularities from the environment. Taken together, it appears that the perceptual crossing paradigm does not yield evidence that unequivocally speaks for the hypothesis that interaction dynamics constitute part of the social cognitive processes that are needed to solve the task.

The conceptual and empirical uncertainties presented above should therefore leave us reluctant to adopt a radical phenactive view and strive to find a less controversial theoretical framework.

## 2.5 If Radicalism Is the Problem, Action-Oriented Predictive Processing Is the Solution

The main problem with radical cognitivism and phenactivism is that they exclude important aspects of social cognition, leaving their depiction of the phenomenon incomplete. While cognitivism does not take into account the importance of interaction and embodiment, it is questionable how phenactive views account for ‘representation-hungry’ elements of social understanding, such as explicit ‘offline’ reasoning about another person. In what follows, I argue that neither radical view can yield a comprehensive view on (social) cognition and that a middle-way is needed.

I previously presented De Bruin and Kästner’s (de Bruin et al. 2012) definitions of cognitivism and (ph)enactivism. They examined which of these theories provide the most comprehensive view of cognition and come to the following conclusion:

To conclude our diagnosis: neither COG [classic cognitivism] nor ENAC [enactive cognition] has been successful in providing a convincing account of both online and offline forms of cognitive processing. It hence seems fruitful to aim at a unified theoretical framework that solves the stalemate between ENAC and COG and integrates online and offline processes into a coherent story of how cognition can best be understood. (de Bruin et al. 2012, p. 547)

What the authors express in this quotation is twofold. First, it shows that there is a *spectrum* of theoretical claims, whose ends appear to be classic cognitivism and (ph)enactivism. Secondly, they rightfully gather that either account is yet to come up with a comprehensive and coherent account of cognition.

The same can be argued for the more specific case of social cognition. Phenactive views have indeed brought to awareness important aspects of the phenomenon that were previously ignored. This mainly concerns the aspects of embodiment, interaction, and the experiential quality of social encounters. Although they have been brought up by traditional phenomenology, they indeed got lost when the philosophical debate focused on cognitivist mind reading schemes. I thus agree with proponents of the phenactive view that a narrow view on the observational inference of mental states does not reflect the manifold nature of social cognition. On the other hand, it is questionable whether phenactive theories are able to capture the whole picture of social cognition. Although they might yield ways to grasp interaction, embodiment, and phenomenology, it is unclear how they would account for other aspects of the phenomenon, such as offline construction of reasons for another person's behavior.<sup>3</sup>

If the goal is to provide a *comprehensive* theoretical framework for social understanding that includes — among others — interaction, it is undesirable to adopt any radical position. As matters stand now, it seems that both cognitivist and phenactive theories have contributed valuable insights to the debate. It could be that some social processes need a rather non-representational, non-computational view, while others require a more cognitivist picture. I therefore argue that we should preserve a middle course and try to prevent any extreme, radical position that potentially excludes important aspects of the phenomenon. It would be advisable to attempt to find a theoretical framework that is able to integrate the full spectrum of social mechanisms.<sup>4</sup> In what follows I suggest that PP yields just the right ideas to do so.<sup>5</sup> In doing so, I will draw on three notions that are central to PP and, or so I shall argue, open up the possibility to combine cognitivist and phenactivist theoretical elements. These three notions are embodied inference (Friston 2012), action-oriented predictions (Clark 2016), and active inference (Friston et al. 2011). I will elaborate on these concepts in the next section, before applying them to the phenomenon of social cognition.

### 3 Action-Oriented Predictive Processing

#### 3.1 Embodied Inference

The notion of ‘embodied inference’ was introduced by Friston and Stephan (Friston and Stephan 2007) to express how PP is an instance of the free-energy principle (FEP). To see what this means, we first have to consider the situation an embodied organism is embedded in. According to the second law of thermodynamics, the entropy of a closed system increases with time. Biological systems, however, are considered open systems in that they exchange energy and matter with their environment. They thusly resist the second law of thermodynamics and sustain their order. How is that achieved? Friston and Stephan (Friston and Stephan 2007, p. 422) suggest that the “premise here is that the environment unfolds in a thermodynamically structured and lawful way and biological systems embed these laws into their anatomy.” In this sense, we can talk about embodied systems as *being* models of the environment they live in (cf. Friston 2012, pp. 89–90), instead of talking about systems that *have* or *build* models of the world. This is what Friston (Friston 2012) calls “embodied inference”. More specifically, this term expresses that the physiology of a system already presupposes the circumstances it lives in — an organism's phenotype determines its possible state space.

- 3 Please note that thus far, the discussion between cognitivism and phenactivism is of an almost fully theoretical nature. This is partly due to the fact that most empirical designs available are based upon cognitivist assumptions and that thus far, phenactivists have only introduced few empirical designs. Thanks to an anonymous reviewer for raising this point.
- 4 This is not to say that one should or could simply combine phenactivism and cognitivism. Due to some metaphysical incompatibilities, a straightforward combination of the two does not come easy. For a detailed discussion of this matter, see Quadt 2015.
- 5 One possible concern at this point is that PP is built on the conviction that cognition is computation. Since most phenactivists reject such a computational view of the mind, this could lead to a rejection of PP by proponents of phenactivism. There are two ways to tackle this worry. First it should be noted that the aim of this paper is to create a *new* position that merely integrates ideas from each side of the theoretical spectrum, but does not aspire to be fully compatible with both. On the other hand, it could be claimed that phenactivism is not obliged to reject computationalism — a topic that needs to be pursued elsewhere. Thanks to an anonymous reviewer for raising this concern.



### 3.2 Action-Oriented Predictions

The topic of representations is one of the most controversial in the debate between cognitivism and phenactivism, which is why the notion of action-oriented predictions is of such high importance. While it is almost impossible to imagine cognitivism without the concept of representations, phenactivism rejects it entirely. It will thus be vital for our goal of finding a middle-way to alleviate the tensions revolving around this topic. One compelling solution has been proposed by Clark, who started to lay out the concept of ‘action-oriented representations’ in his earlier work (Clark 1997) and continued to draw a picture of representations that defies the ‘old-school’ version of cognitivism. The concept of representation in radical cognitivism refers to (sub-)symbolic vehicles that carry a specific content and in this sense are thought to ‘mirror’ the external world. Clark offers several arguments in favor of the idea that the kind of representation that PP yields is in no way related to the stiff, passive-mirror-of-nature representation old-fashioned cognitive science talked about.

First, although internal models are a central part of PP, these models are fundamentally grounded in embodiment, in that they “allow a system to combine a real sensorimotor grip on dealing with its world with the emergence of higher-level abstractions that (crucially) develop in tandem with that grip.” (Clark 2014, p. 242) Representations or internal models are not marooned from brain-external matter, they are *for* engaging the body and world, to elicit action and active navigation of the environment. At the same time, the concept of representations is not given up. To see how representations in this context are defined, allow me to cite Clark’s idea at length:

[...] each PP level (perhaps these correspond to cortical columns — this is an open question) treats activity at the level below as if it were sensory data, and learns compressed methods to predict those unfolding patterns. This results in a very natural extraction of nested structure in the causes of the input signal, as different levels are progressively exposed to different re-codings, and re-re-codings of the original sensory information. These re-recodings [...] enable us, as agents, to lock us onto worldly causes that are ever more recondite, capturing regularities visible only in patterns spread over space and time. Patterns such as weather fronts, persons, elections, marriages, promises, and soccer games. [...] What locks the agent on to these familiar patterns is, however, the whole multi-level processing device (sometimes, it is the whole machine in action). That machine works (if PP is correct) because each level is driven to try to find a compressed way to predict activity at the level below, all the way out to the sensory peripheries. These nested compressions, discovered and annealed in the furnace of action, are what I [...] would like to call “internal representations. (Clark 2015a, p. 5)

As I read Clark, the essence of his claim is that representations are abstractions of sensory signals. They are not the sensory data themselves, but carry information that has been compressed and abstracted, enabling a prediction of what the “sensory” data a level below could be. In this sense, it is useful to talk about internal models and representations. Predictions represent potential sensory input, becoming more and more abstract as one goes up the hierarchy.

These kinds of representations do not merely generate a picture of the world in our heads. If the central role active inference plays in FEP is taken seriously, representations engage the *whole agent* to extract hidden causes in the world. In this sense, Clark opts for talking about ‘action-oriented’-predictions: “They will represent how things are in a way that, once suitably modulated by the precision-weighting of prediction error, also prescribes (in virtue of the flows of sensation they predict) how to act and respond.” (Clark 2016, p. 133). Considering the role of internal models—to prepare systems to act upon their environment and enable them to do so—thus helps us tune the notion of representation towards a more embodied, flexible one. This is, in my view, a crucial step towards

finding the ‘golden middle’ between cognitivist and enactive theories. The notion of action-oriented predictions will also be of high importance for what I call ‘interactive inference.’

### 3.3 Active Inference

Clark’s interpretation of Friston’s take on FEP entails that organisms strive to reduce free energy by opting for the most efficient way to do so. Efficiency, here, refers to finding the strategy that involves the least complex route towards prediction error minimization, while bringing the largest effect. The brain’s task thus not only entails the construction of inner models, but also preparing an organism for its bodily exchange with the environment. This involves the estimation of which channel and which ‘strategy’ will most efficiently minimize prediction error — will it be better to change my models (perception) or use my body to bring forth a change in the environment (action)? The latter strategy refers to what is called ‘active inference.’

The body thus has an indispensable role in action-oriented PP. As described in the previous section, the trick is to acknowledge that the task of predictive models (i.e., representations) is to find the most efficient, least costly route to success. This is what Clark (Clark 2015a, p. 9) refers to when he talks about the “productive laziness” of the brain; whenever the body or the world can be recruited to do a job, there is no need to compute complex inner models. Precision-weighting determines whether low-level modalities or high-level modeling will ‘be in charge’ to solve the task at hand — depending on how efficient the strategy is estimated to be. This strategy will more often than not involve the engagement of brain-external structures:

The task of the generative model [...] is to capture the simplest approximation that will support the actions required to do the job — this means taking into account whatever work can be done by a creature’s morphology, physical actions, and socio-technological surroundings. [...] There is thus no conflict with work that stresses biological frugality, satisficing, or the ubiquity of simple but adequate solutions that make the most of brain, body, and world. (Clark 2015a, p. 291)

Clark here endorses a central aspect of phenactive theories, namely the role of extra-neural structures for an agent’s navigation of its environment. Active inference takes center stage in this interpretation of PP, in virtue of the fact that the function of predictive models is to distribute the cognitive workload and recruit embodied action whenever possible.

Such a view emphasizes that PP displays a deep and fundamental connection of mind and body. This leaves us with the following picture of the (social) mind. PP accounts for quite a spectrum of phenomena; on the one hand, it is a rather brain-bound view, since the generation of predictions and the precision weighting process is neurally implemented. In that way, perception is brought forth mainly by top-down processing and is determined internally. This side of PP neatly accommodates ‘representation-hungry’ processes like imagination, dreaming and also thinking about other people, which seem to occur without much brain-external help. To see this, consider that it is argued that the main task of the cortex is to generate predictions about incoming stimuli (Friston et al. 2012). This means, basically, that the brain is able to reconstruct “the sensory signal using knowledge about interacting causes in the world” (Clark 2016, p. 85). Once learned, the system will be able to process without actual input and thus bring forth imagination, dreams, or explicit theorizing.

On the other hand, even those more ‘decoupled’ phenomena have been shown to involve the body. Saccadic eye-movements, for example, may be the bodily ‘grounds’ for phenomenal experience in dream states (Metzinger 2014). In this sense, the body and environment are indispensable parts of cognition. It is this neat interplay of internal models, action, and the body that make PP the perfect fit for a theory that integrates both phenactive and cognitivist elements, providing a sound ground for a theory on social cognition.

## 4 Conceptual Tools

### 4.1 Embodied Social Inference (EmSI)

The first concept I wish to introduce is ‘embodied social inference’ (EmSI), which emphasizes that the physiology of an organism constrains the kinds of social interactions it can engage in. Recall that embodied inference means that the thermodynamical laws of an agent’s environment are ‘folded into’ her morphology; that her body is built to keep her alive by resisting the second law of thermodynamics. In this sense, it can be said that the agent is a model of its world, because their physiology incorporates the physical laws the body needs to obey to ensure survival. This is related to the claim that the physiology of an organism constrains the kind of mind it has, because the laws that are relevant for this specific phenotype will be modeled by its body.

In the same way, it can be said that the kind of body an organism has determines the kind of social interaction and understanding it is capable of. While a herring strives to stay in its large fish school to ensure its survival, cats aim for much smaller groups or may even survive on their own. The human body needs a caretaker for an extended amount of time during childhood, not being able to sustain itself until a certain age. Further, while humans are able to use their speech-apparatus to communicate and interact, ants will have to rely on pheromones to send signals to each other. This can be seen as embodied social inference (EmSI); an organism’s phenotype determines the kind of social abilities they possess. To be more specific, an embodied organism can be called a model of its social environment because their physiology incorporates possibilities for interaction; vocal cords make vocal communication possible, for example.

This is also important when discussing the role of similarity for social cognition. While there are very many individual differences, the gross anatomy and morphology of individual organisms of one species is rather similar. This similarity may provide a fundamental role in the attempt to recognize the other as ‘one of us’ and thus to understand them. The role of bodily similarity is twofold; it not only determines how well we understand another person, but it also opens up the possibility that there needs to be a general similarity for social processing to begin with. The claim that a certain degree of similarity is needed in order to understand each other has been famously formulated by a number of researchers. For example, Meltzoff (e.g., [Meltzoff 2005](#); [Meltzoff 2007](#); [Meltzoff 2013](#)) states in his ‘like me’ hypothesis that the development of understanding others hinges upon the fact that the infant perceives the other as ‘like me’. In fact, it is claimed

that the core sense of similarity to others is not the culmination of social development, but the precondition for it. Without this initial felt connection to others, human social cognition would not take the distinctively human form that it does. ([Meltzoff 2013](#), p. 139)

Meltzoff’s reasoning rests on the assumption that social cognition — especially in developmental terms — is enabled by matching visual to motor representations. The bedrock of his argument rests on many neonatal imitation studies by him and his colleagues ([Meltzoff and Moore 1997](#)). Although having no visual information about one’s own face, newborn babies appear to be able to imitate an adult’s behavior, such as tongue protrusion ([Meltzoff and Moore 1977](#)). It is thought that the visual information of the adult is ‘matched’ onto the proprioceptive information the newborn already acquired. This matching process then enables imitative behavior.

The ‘like me’ hypothesis gains additional support when viewed from a PP perspective. In accordance with a simulation model, Friston and Frith ([Friston and Frith in press](#), p. 12) argue that “internal or generative models used to infer one’s own behaviour can be deployed to infer the beliefs (e.g., intentions) of another — provided both parties have sufficiently similar generative models.” In other words, similarity here is seen as a presupposition for mental state inference. Only when there is a suffi-

cient similarity of models, there can also be a big enough overlap which allows the application of one's own models to understand the other's behavior.<sup>6</sup>

This is important for several reasons. First, I claim that replicative interactive inference (RInI) largely draws on similarity. Secondly, similarity relates to the discussion of so-called 'shared representations'. What Friston and Frith refer to above is exactly this — models or representations that are sufficiently similar can be used for both self- and other-related processing. A famous example of social mechanisms that rely on shared representations is found in the mirror neuron system. Mirror neurons are known to fire not only when an individual executes, but also when she merely observes an action (e.g., [Rizzolatti and Craighero 2004](#)). They can thus be said to involve shared representations, because they function both for action execution (self-related) and action observation (other-related). Finally, action-oriented PP implies that representations, i.e., predictive models, are grounded in sensorimotor processes. The range of these processes, in turn, are constrained by the kind of body an organism has. As trivial as it seems, this basically means that bodies determine the range of (social) experiences one can have. Metzinger ([Metzinger 2004\[2003\]](#), pp. 160–161) picks up this point and formulates it as the 'single-embodiment constraint':

Trivially, the causal interaction domain of physical beings is usually centered as well, because the sensors and effectors of such beings are usually concentrated within a certain region of physical space and are of limited reach. [...] This functional constraint is so general and obvious that it is frequently ignored: in human beings, and in all conscious systems we currently know, sensory and motor systems are physically integrated within the body of a single organism. This singular "embodiment constraint" closely locates all our sensors and effectors in a very small region of physical space, simultaneously establishing dense causal coupling.

In making this statement, Metzinger clarifies that the behavioral space of an individual is limited and constrained by its body. The range of possible behavior and experiences shape our cognitive processing, an effect whose pervasiveness becomes clear when viewed through the lens of PP. PP depicts the neural and cognitive architecture as immensely flexible and ever-changing. If precision-weighting admits, any sensory signal can change predictions at any level of the processing hierarchy.

If it is furthermore true that anatomical as well as morphological features are the basis for a system's generative models, and if it is true that these models can only be used for both self- and other-related processing if they are sufficiently similar, it follows that the bodies of interacting individuals must be sufficiently similar, too. Put differently, if the bodily structure of individuals is grossly different, their models may not be sufficiently similar, thus restricting interaction and understanding. The relation to EmSI should be clear by now; the phenotype of an individual must exhibit some degree of similarity in order to make it possible to recognize others as 'like me' and thus to enable the matching of one's own models to the other's.

To sum up, EmSI refers to the determining and constricting role that bodies play for social cognition, and also for interaction. In this sense, it can be said that the very physiology of an individual determines its space of possible social interactions.

## 4.2 Interactive Inference

The notion of interactive inference is tightly related to active inference and can be described as the minimization of prediction error by engaging in an embodied interaction. Applying this line of thought to the realm of social cognition, I now wish to add that interaction can play the same role for social cognitive processing as action plays for general cognition. This amounts to gathering information about

<sup>6</sup> While similarity is claimed to be crucial for social cognition, please note that it may not be necessary for all kinds of social understanding. Otherwise, we would not be able to understand that the dog's wagging tail is an expression of his excitement or that what the octopus is intending is to open the jar. Distinguishing between self and other, thus, is just as important as similarity. Thanks to an anonymous reviewer for raising this concern.

the social environment and in this way actively sculpting one's external and internal environment. I thus claim that just as *active inference* is central for general cognition, interactive inference (InI), as I will call the process, is as central for social cognition.

What exactly does 'interactive inference' mean? Recall that active inference can be described as minimizing prediction error in several ways, namely by actively changing an agent's inner and outer environment so to fulfill exteroceptive, proprioceptive and interoceptive predictions, and the disambiguation between competing predictive models (cf. Seth 2015, pp. 13–14). In a similar way, interactive inference can be described as minimizing prediction error while navigating the social environment. Instead of changing one's model about the other person in order to understand her (perceptual inference), InI serves to actively sample proof for predictions or to cancel out possible models about causes of the behavior or another person. The basic idea is that engaging in interactions with other people can be a means to minimize prediction error and thus offers a fast and fruitful way to understand others. In what follows, I will elaborate on the concept by further distinguishing two types of InI; replicative interactive inference (RInI) and complementary interactive inference (CInI). The distinction serves to distill and differentiate the manifold ways in which interaction can enrich and enable social cognitive processing.

### 4.3 Replicative Interactive Inference (RInI)

Turning towards two different types of InI, let us first consider what I will call replicative interactive inference (RInI). In RInI, the other's internal or bodily states are replicated, such as in mimicry, emotional contagion, or automatic imitation. This replication has two effects, both of which can be said to make prediction error minimization more efficient. First, it serves to make oneself more similar to the other; in other words, to 'put oneself into' the other's bodily state. Instead of generating brand new models about the other person and the possible causes of her behavior on the basis of exteroceptive social stimuli, it will be quicker to gather information by tuning into their bodily, i.e., interoceptive or proprioceptive, state. So far, we have discussed the role of similarity in terms of morphology. However, this referred to the basic possibility of understanding each other. RInI can now be said to enhance this similarity by replicating the other's *current* bodily state.

Secondly, RInI serves to give 'first-hand' information about the other person. In order to get a sense of the other, predictions about their current state are corrected in virtue of error signals. When replicating the other's bodily state, these error signals should be more reliable, since they come not only from one exteroceptive (e.g. visual) source, but also from an internal source (e.g., proprioceptive prediction error). Therefore, during RInI, the bodily state (e.g., posture, movements) of another person is mimicked in order to supplement exteroceptive information about them with interoceptive and proprioceptive information. Mimicry, synchronization and automatic imitation are instances of RInI that function to make predictions about the other more precise by increasing the number of signal sources that yield relevant information.

These phenomena occur automatically and involuntarily — even when people are explicitly asked to suppress these tendencies. There are, for example, many studies which show that individuals cannot help but synchronize their movements with the other person. This has been shown for several motor acts, such as finger tapping (Oullier et al. 2008), rocking in rocking chairs (Richardson et al. 2007) and body posture (Lafrance and Broadbent 1976). Chartrand and Lakin (Chartrand and Lakin 2013, p. 288) provide a comprehensive review on these effects and summarize them under the notion of 'the Chameleon effect': "[...] much like chameleons change their color to blend into their surrounding environment, humans alter their behavior to blend into their social environment." In a vast number of studies that are reviewed by the authors, it has been shown that mimicry and synchronization are accompanied by many facilitating factors and in turn also facilitate social interaction. For example, individuals are more likely to mimic another person when there are prior 'pro-social' factors, such as

in-group effects and prior rapport. Individuals with similar opinions and high empathy rates are more prone to mimicry and synchronization. Although there are also inhibitors of mimicry such as the wish to disaffiliate with the other person, the authors conclude that unconscious mimicry and synchronization seems to be a default for social interactions and occurs even when individuals face other tasks (cf. [Chartrand and Lakin 2013](#), p. 290). Further, individuals that were told to keep still and suppress their tendency to replicate the other person's behavior perform worse at emotion detection tasks.

Furthermore, consider the following study conducted by Ainley and colleagues ([Ainley et al. 2014](#)) that links interoceptive awareness with the tendency to automatically imitate. They found that — contrary to their initial prediction — participants who scored higher for interoceptive awareness had a greater tendency to imitate. In other words, the more one is aware of one's interoceptive processing (in this study measured with the so-called 'heartbeat perception task'), the less one is able to inhibit automatic imitation. One possible (although rather speculative) interpretation of these results is that people with higher interoceptive awareness set the gain on interoceptive prediction errors higher. Ainley and colleagues ([Ainley et al. 2014](#), p. 26) hypothesize that

[g]iven that interoceptive awareness affects perception of the body, it is also likely to modulate action representations. It has recently been indicated that in order to avoid mirroring another person's actions it is essential to reduce the precision of proprioceptive prediction error (Friston, Mattout & Kilner, 2011). If people with high interoceptive awareness have initially precise proprioceptive prediction errors then their tendency to imitate others may be accounted for.

Put differently, in order to inhibit imitation and not to replicate the other's movement, gain on prediction error must be set low. Thus, weighting the precision of prediction errors high may result in the tendency to automatically imitate the other person. If this is correct, the processing steps underlying automatic imitation could be the following. First, contextual cues yield information that the current incoming signals originate from another person; thus representations about sensory consequences — which could be proprioceptive, interoceptive, or exteroceptive — are recruited. Next, depending on whether the gain on prediction error is set high or low, the observed state of the other person is replicated or not. As described above, highly precise errors would result in a replication of the other's state, while low-weighted prediction errors would result in the inhibition of automatic imitation.

This may not only be the case in motor imitation. Phenomena such as emotional contagion or the queasiness one feels when observing someone eating something truly disgusting could be cases in which gain on interoceptive prediction error is set high. This would lead to the replication of the other's interoceptive state and thus trigger 'shared bodily experiences'. Entering an actual interaction should provide all interacting individuals with more unambiguous cues to which predictive model has the highest posterior probability. To see this, recall that RInI serves to make the bodies of interacting individuals to be in more similar states. If it is true that higher-order representations are grounded in sensorimotor processes, this should also lead to a more similar representational state of the body model in both individuals.

Facial emotion recognition serves as another elaborative example of RInI. Several findings are of central importance here. First, it has been claimed that the face is likely the most significant body region for social cognition, since it provides the most relevant information when it comes to understanding others (cf. [Farmer et al. 2014](#), p. 290). Secondly, a great number of studies have shown that the sight of emotional expressions leads to activation in brain areas with mirror properties (e.g., [Wicker et al. 2003](#)). Further, people tend to mimic facial expressions of their interaction partners (cf. [Chartrand and Lakin 2013](#), p. 287). Above, I reviewed some of the research suggesting that mimicry not only occurs ubiquitously, but that it also has striking effects on social relationships. In turn, there is growing evidence that the tendency to mimic is considerably influenced by top-down effects and prior information about the other person (*ibid.*).

Putting these findings together, the following picture emerges: Visual signals of the other person's facial expression (plus contextual information) trigger generative models about the underlying emotional state — this is where shared representations enter the picture. These predictive models serve as a basis for generating proprioceptive predictions — that is, the motor commands underlying the facial expression — and also interoceptive predictions which refer to the internal bodily state the person must have been in to give rise to the emotion displayed on their face. Proprioceptive prediction error can be quashed by changing the state of facial muscles ourselves, thus mimicking the other person. Interoceptive prediction error can also be minimized by actively changing one's internal environment. Seth (Seth 2013) claims that emotions occur when prediction errors are cancelled out for exteroception, interoception and proprioception, thus disambiguating multimodal models generated in the insular cortex. The same may be true for emotion recognition; multimodal predictive models about the cause of incoming exteroceptive signals are confirmed or ruled out by quashing proprioceptive and interoceptive prediction error, inferring the most likely cause of the observed emotion. Mimicry, as an instance of RInI, is therefore a crucial and fast way to enhance this process of emotion recognition.

The rationale here is that greater bodily similarity will lead to greater social similarity and facilitate social understanding. Of course, whether or not interactive inference will be deemed a fruitful way to figure out the other person depends on prior beliefs and expectations about the other person. As already mentioned, top-down effects are pervasive and determine whether or not mimicry occurs. However, this fits nicely in the more general framework of PP, since the multidirectional interplay between bottom-up and top-down effects is of central importance.

#### 4.4 Complementary Interactive Inference

The automatic replication of bodily and motor states is, of course, not the only process which happens between individuals during an interaction. Instead of replicating, it will often be necessary to perform complementary actions. This second case I will call complementary interactive inference (CInI). CInI refers to changing one's internal (i.e., bodily) or external environment in response to the other person without replicating the other's state. This has several functions.

First, it can serve to regulate another person's current bodily or emotional state. This can be achieved by changing one's own posture, movements, or gestures (e.g., giving the other an encouraging nod to make her continue talking), but also by altering one's interoceptive state. An intriguing example of this latter case can be found in so-called 'kangaroo care', which is often used for prematurely born (human) babies. During kangaroo care, mothers hold their infants in an upright position close to their body between their breasts and underneath their clothing. It has been found that this has many positive effects on both mother and baby. Most interestingly for the matter here are the physiological effects; mothers regulate their body temperature according to their infants needs and thereby also enhance self-regulation of the child. When the child has a fever, mothers lower their body temperature so to provide cooling for their infant. Further, if the baby has an irregular heartbeat, this can be counteracted and becomes more steady when their ear is placed on their mother's chest and they hear the mother's steady heartbeat (Ludington-Hoe et al. 2006; Nyqvist et al. 2010).

A second function of CInI could be to evoke behavioral response of the other person that serves as additional exteroceptive input in order to disambiguate social stimuli. Gestures, facial expressions or other movements are used to signal one's uncertainty and thus provoke a reaction of the other person, which then serves as additional information. I might, for example, shrug my shoulders or raise my eyebrows in order to signal you that I did not understand what you were saying. This signals to you, in turn, that I need additional information and may — if this interaction is successful — elaborate on your stance.

Thirdly, CInI serves to make oneself more predictable, thereby smoothing out social understanding, joint action, or coordination. For joint actions that require coordination, for example, Vesper and

colleagues (Vesper et al. 2010) have coined the term ‘coordination smoothers’ to describe the modulation of one’s own behavior in order to make coordination with another person more simple:

One way to facilitate coordination is for an agent to modify her own behavior in such a way as to make it easier for others to predict upcoming actions, for example by exaggerating her movements or by reducing the variability of her actions. (Vesper et al. 2010, p. 999)

In several studies it has been found that people indeed adjust their movement trajectories, their pace or use signaling or communicative actions in order to increase predictability. For example, piano players that are performing a duet exaggerate their finger movements or speed up in order to decrease variability (Keller et al. 2007).

What may be the mechanisms underlying all these functions of CInI? Vesper and colleagues claim that prediction and motor simulation are key to enabling the execution of complementary actions. Simulations are thought to be especially useful for joint actions, where they enhance timing and anticipation of sensory consequences. This comes naturally within a predictive processing framework, since it is assumed that predictive models represent sensory consequences of actions in a counterfactual manner (Seth 2014). Assuming that these predictive models can be shared — i.e., that they can be used for both self- and other-related processing — it becomes clear how they can be exploited to compute not only the consequences of one’s own, but also the other’s sensorimotor trajectory. Interestingly, the role of similarity becomes important one more time, for joint action is enhanced when the timing patterns of both agents are predictable. The predictability in turn is dependent upon how similar agents are, and how similar their motor experience is. This has been shown in several studies that show that mirror neuron activity increases when observing actions that already belong to one’s own motor repertoire (Calvo-Merino et al. 2004). The mirror neuron system is therefore involved in both replicative action processing and the preparation of complementary actions. According to Pezzulo and colleagues (Pezzulo and Dindo 2011, p. 612), “this suggests that the brain can encode actions executed by others in an interaction-oriented way, and more broadly that action-perception mappings could be quite flexible and task-dependent.”

Taken together it can thus be hypothesized that shared predictive models are not only useful for replicative, but also complementary interactive inference. Interaction is here used to solve problems with the other person, in virtue of making oneself more predictable, and using one’s body to signal what is needed from the other. Framing interaction within PP allows to attribute an important role to interaction patterns between individuals to their social cognitive processing. The mutually unfolding predictions, actions, counteractions and perceptions are captioned by interactive inference and thus provide a new way to conceptually grasp how interactions matter for social cognition.

## 5 Conclusion

The current situation in the research field of social cognition has been depicted as problematic because the theoretical schemes of phenactivism and cognitivism alone do not yield a sound ground for a theoretical framework on the phenomenon. While the latter ignores the importance of embodied interaction, the former has been doubted to have sufficient conceptual and empirical back-up. At the same time, both theories account for important aspects of social cognition. The main goal of this paper was therefore to find a theoretical approach to combining these aspects, while circumventing the problems that come with phenactivism and cognitivism.

Action-oriented PP provides many opportunities for implementing both cognitivist and phenactivist elements in a theory on social cognition. Another aim in this paper was thus to exploit some of them and start to suggest ways in which PP can enlighten theoretical work on the phenomenon. Just like general cognition, social cognition heavily draws on the interaction of body, mind and world. PP



is therefore the perfect partner to highlight this dependency, since it appears that although a great part of the prediction error minimization machine is located in the brain, the body and action play an indispensable role for this mechanism. To see this, remember that while prediction generation clearly is the brain's job, the minimization of prediction error — the core of PP — heavily engages the body and world in virtue of active inference.

In this sense, it has been claimed that embodiment is fundamental to (social) cognition in at least two ways. First, the very morphology and phenotype of a system set the baseline of what are probable states for it to be in. To capture this idea for the social realm, the notion of embodied social inference (EmSI) has been introduced. EmSI expresses that our bodies define the kinds of social interactions we are able to engage in, and that a certain amount of morphological similarity is needed in order to enable social understanding. Second, as described above, active inference appears as a part of PP it cannot do without. This has consequences for our view on both general and social cognition. Concerning the latter, I coined the term of 'interactive inference' (InI) to describe replicative and complementary behavior that serves to cancel out prediction error via engagement in social interactions.

The perspective adopted in this paper has implications for future research. For example, it is asserted that differences in sensorimotor processes result in differences of predictive models, which can be shared and exploited for social cognition. From this, we can derive the prediction that large differences of sensorimotor processes between individuals will make social cognitive processes that rely on them more difficult. This has been shown for the case of autism. Cook (Cook 2016) argues that since the kinematics of movements in typical and autistic individuals deviate, they are less likely to resonate with each other. This may be one cause not only for the social impairments that come with autism, but also for the difficulties that typical individuals have in understanding autistic individuals. From the perspective of interactive inference, it can be assumed that processes of replication are disrupted, thus leading to an impaired inference process between individuals. It can be hypothesized that predictive models that are built on the basis of an individual's own motor repertoire are too different to support a stable inference mechanism. Future research should investigate at which level impairments occur and cause impaired social interactions between autistic and neurotypical individuals.

This also relates to the question of how individual differences influence social cognition. This question needs to be broken down into several sub-issues and more attention in future research. As described before, in the case of autism it has been hypothesized that differences in kinematic profiles between individuals with autism and typically developed individuals are one source of problems in social understanding (Cook 2016). It has further been shown that similar motor experience of individual enhances imitative behavior (Kilner et al. 2007). These findings can serve as a starting point to examine how important individual similarity and differences are for social cognitive processing. At the neural level, differences in precision weighting could influence the tendency to imitate. This would be predicted by the claim that precision optimization is a leading component in automatic imitation.

The considerations in this paper show that although PP puts forth a quite central role of the brain, it still integrates a deep sense of embodiment and relation with the environment in virtue of being an instance of FEP. As such, this theory displays a fundamental continuity of mind and life. Again, there lies a great opportunity to satisfy demands from phenactivism in taking this continuity seriously and explore its consequences for a theory of our social minds.

## References

- Ainley, V., Brass, M. & Tsakiris, M. (2014). Heartfelt imitation: High interoceptive awareness is linked to greater automatic imitation. *Neuropsychologia*, *60*, 21–28. <https://dx.doi.org/10.1016/j.neuropsychologia.2014.05.010>.
- Arnau, E., Estany, A., González del Solar, R. & Sturm, T. (2014). The extended cognition thesis: Its significance for the philosophy of (cognitive) science. *Philosophical Psychology*, *27* (1), 1–18. <https://dx.doi.org/10.1080/09515089.2013.836081>.
- Auvray, M. & Rohde, M. (2012). Perceptual crossing: The simplest online paradigm. *Frontiers in Human Neuroscience*, *6*. <https://dx.doi.org/10.3389/fnhum.2012.00181>.
- Auvray, M., Lenay, C. & Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, *27* (1), 32–47. <https://dx.doi.org/10.1016/j.newideapsych.2007.12.002>.
- Bick, J., Zhu, T., Stamoulis, C., Fox, N., Zeanah, C. H. & Nelson, C. A. (2015). Effect of early institutionalization and foster care on long-term white matter development: A randomized clinical trial. *JAMA Pediatrics*. <https://dx.doi.org/10.1001/jamapediatrics.2014.3212>.
- Calvo-Merino, B., Glaser, D. E., Grézes, J., Passingham, R. E. & Haggard, P. (2004). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, *15* (8), 1243–1249. <https://dx.doi.org/10.1093/cercor/bhi007>.
- Chartrand, T. L. & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology*, *64*, 285–308. <https://dx.doi.org/10.1146/annurev-psych-113011-143754>.
- Clark, A. (1997). The dynamical challenge. *Cognitive Science*, *21* (4), 462–481.
- (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–253. <https://dx.doi.org/10.1017/S0140525X12000477>.
- (2014). Perceiving as predicting. In D. Stokes, M. Mohan & S. Biggs (Eds.) *Perception and its modalities* (pp. 23–44). Oxford: Oxford University Press.
- (2015a). Embodied prediction. In T. K. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt am Main: MIND Group. <https://dx.doi.org/10.15502/9783958570115>.
- (2015b). Predicting peace: The end of the representation wars. In T. K. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt am Main: MIND Group. <https://dx.doi.org/10.15502/9783958570979>.
- (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. New York, NY: Oxford University Press.
- Cook, J. (2016). From movement kinematics to social cognition: The case of autism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371* (1693). <https://dx.doi.org/10.1098/rstb.2015.0372>.
- de Bruin, L., Van Elk, M. & Newen, A. (2012). Reconceptualizing second-person interaction. *Frontiers in Human Neuroscience*, *6*. <https://dx.doi.org/10.3389/fnhum.2012.00151>.
- De Jaegher, H. & Di Paolo, E. A. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences* (6), 485–507. <https://dx.doi.org/10.1007/s11097-007-9076-9>.
- De Jaegher, H., Di Paolo, E. A. & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14* (10), 441–447. <https://dx.doi.org/10.1016/j.tics.2010.06.009>.
- Di Paolo, E. A. & De Jaegher, H. (2012). The interactive brain hypothesis. *Frontiers in Human Neuroscience*, *6*, 1–16. <https://dx.doi.org/10.3389/fnhum.2012.00163>.
- Farmer, H., McKay, R. & Tsakiris, M. (2014). Trust in me: Trustworthy others are seen as more physically similar to the self. *Psychological Science*, *25* (1), 290–292. <https://dx.doi.org/10.1177/0956797613494852>.
- Fox, N., Almas, A. N., Degnan, K. A., Nelson, C. A. & Zeanah, C. H. (2011). The effects of severe psychosocial deprivation and foster care intervention on cognitive development at 8 years of age: Findings from the Bucharest Early Intervention Project. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *52* (9), 919–928. <https://dx.doi.org/10.1111/j.1469-7610.2010.02355.x>.
- Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13* (7), 293–301. <https://dx.doi.org/10.1016/j.tics.2009.04.005>.
- (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11* (2), 127–138. <https://dx.doi.org/10.1038/nrn2787>.
- (2012). Embodied inference: Or “I think therefore I am, if I am what I think”. In J. Kriz (Ed.) *The implications of embodiment: Cognition and communication* (pp. 89–125).
- Friston, K. & Frith, C. (in press). A duet for one. *Consciousness and Cognition*. <https://dx.doi.org/10.1016/j.con-cog.2014.12.003>.

- Friston, K. J. & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159 (3), 417–458. <https://dx.doi.org/10.1007/s11229-007-9237-y>.
- Friston, K., Mattout, J. & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, 104 (1-2), 137–160. <https://dx.doi.org/10.1007/s00422-011-0424-z>.
- Friston, K., Adams, R. A., Perrinet, L. & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, 3, 151. <https://dx.doi.org/10.3389/fpsyg.2012.00151>.
- Froese, T., Iizuka, H. & Ikegami, T. (2014). Embodied social interaction constitutes social cognition in pairs of humans: A minimalist virtual reality experiment. *Scientific Reports*, 4. <https://dx.doi.org/10.1038/srep03672>.
- Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17 (2), 535–543. <https://dx.doi.org/10.1016/j.concog.2008.03.003>.
- Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2 (12), 493–501.
- Gopnik, A. & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, 7 (1-2), 145–171. <https://dx.doi.org/10.1111/j.1468-0017.1992.tb00202.x>.
- Herschbach, M. (2012). On the role of social interaction in social cognition: A mechanistic alternative to enactivism. *Phenomenology and the Cognitive Sciences*, 11, 467–486. <https://dx.doi.org/10.1007/s11097-011-9209-z>.
- Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C. & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, 3 (3), 529–535. <https://dx.doi.org/10.1371/journal.pbio.0030079>.
- Keller, P. E., Knoblich, G. & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, 16 (1), 102–111. <https://dx.doi.org/10.1016/j.concog.2005.12.004>.
- Kilner, J., Friston, K. & Frith, C. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, 8 (3), 159–166. <https://dx.doi.org/10.1007/s10339-007-0170-2>.
- Lafrance, M. & Broadbent, M. (1976). Group rapport: Posture sharing as a nonverbal indicator. *Group & Organization Management*, 1 (3), 328–333. <https://dx.doi.org/10.1177/105960117600100307>.
- Ludington-Hoe, S. M., Lewis, T., Morgan, K., Cong, X., Anderson, L. & Reese, S. (2006). Breast and infant temperatures with twins during shared kangaroo care. *Journal of Obstetric, Gynecologic, and Neonatal Nursing: JOGNN / NAACOG*, 35 (2), 223–231. <https://dx.doi.org/10.1111/j.1552-6909.2006.00024.x>.
- Meltzoff, A. N. (2005). Imitation and other minds: The “like me” hypothesis. In S. Hurley (Ed.) *Perspectives on imitation* (pp. 55–77). Cambridge, MA: MIT Press.
- (2007). The ‘like me’ framework for recognizing and becoming an intentional agent. *Acta Psychologica*, 124 (1), 26–43. <https://dx.doi.org/10.1016/j.actpsy.2006.09.005>.
- (2013). Origins of social cognition. In M. R. Banaji & S. A. Gelman (Eds.) *Navigating the social world* (pp. 139–144). Oxford University Press. <https://dx.doi.org/10.1093/acprof:oso/9780199890712.003.0025>.
- Meltzoff, A. N. & Moore, K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 75–78.
- (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting*, 6, 179–192.
- Metzinger, T. (2004[2003]). *Being no one: The self-model theory of subjectivity*. Cambridge, MA: MIT Press.
- (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). Routledge.
- Nyqvist, K. H., Anderson, G. C., Bergman, N., Cattaneo, A., Charpak, N., Davanzo, R., Ewald, U., Ibe, O., Ludington-Hoe, S., Mendoza, S., Pallás-Allonso, C., Ruiz Peláez, J. G., Sizun, J. & Widström, A.-M. (2010). Towards universal kangaroo mother care: Recommendations and report from the first European conference and seventh international workshop on kangaroo mother care. *Acta Paediatrica*, 99 (6), 820–826. <https://dx.doi.org/10.1111/j.1651-2227.2010.01787.x>.
- Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J. & Kelso, J. A. S. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience*, 3 (2), 178–192. <https://dx.doi.org/10.1080/17470910701563392>.
- Overgaard, S. & Michael, J. (2013). The interactive turn in social cognition research: A critique. *Philosophical Psychology*, 1–25. <https://dx.doi.org/10.1080/09515089.2013.827109>.
- Pezzulo, G. & Dindo, H. (2011). What should I do next? Using shared representations to solve interaction problems. *Experimental Brain Research*, 211 (3-4), 613–630. <https://dx.doi.org/10.1007/s00221-011-2712-1>.

- Quadt, L. (2015). Multiplicity needs coherence – Towards a unifying framework for social understanding. In T. K. Metzinger & J. M. Windt (Eds.) *Open MIND*: 26(C). Frankfurt am Main: MIND Group. <https://dx.doi.org/10.15502/9783958571112>.
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L. & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26 (6), 867–891. <https://dx.doi.org/10.1016/j.humov.2007.07.002>.
- Rizzolatti, G. & Craighero, L. (2004). The mirror neuron system. *Annual Review of Neuroscience*, 27 (1), 169–192. <https://dx.doi.org/10.1146/annurev.neuro.27.070203.144230>.
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17 (11), 565–573. <https://dx.doi.org/10.1016/j.tics.2013.09.007>.
- (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, 5 (2), 97–118. <https://dx.doi.org/10.1080/17588928.2013.877880>.
- (2015). The cybernetic Bayesian brain. In T. K. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt am Main: MIND Group. <https://dx.doi.org/10.15502/9783958570108>.
- Thompson, E. (2010). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, Mass and London: Belknap.
- Varela, F. J., Rosch, E. & Thompson, E. (1993). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.
- Vesper, C., Butterfill, S., Knoblich, G. & Sebanz, N. (2010). A minimal architecture for joint action. *Neural networks: The Official Journal of the International Neural Network Society*, 23 (8-9), 998–1003. <https://dx.doi.org/10.1016/j.neunet.2010.06.002>.
- Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V. & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, 40 (3), 655–664. [https://dx.doi.org/10.1016/S0896-6273\(03\)00679-2](https://dx.doi.org/10.1016/S0896-6273(03)00679-2).
- Zimmer, D. E. (1989). *Wilde Kinder*. In D. E. Zimmer (Ed.) *Experimente des Lebens* (pp. 21–47). Zürich: Haffmanns Verlag.