Phenom Cogn Sci (2019) 18:401–424
https://doi.org/10.1007/s11097-018-9577-8

Mechanistic explanation for enactive sociality

Ekaterina Abramova[1] · **Marc Slors**[1]

Published online: 7 June 2018

Abstract In this article we analyze the methodological commitments of a radical embodied cognition (REC) approach to social interaction and social cognition, specifically with respect to the explanatory framework it adopts. According to many representatives of REC, such as enactivists and the proponents of dynamical and ecological psychology, sociality is to be explained by (1) focusing on the social unit rather than the individuals that comprise it and (2) establishing the regularities that hold on this level rather than modeling the sub-personal mechanisms that could be said to underlie social phenomena. We point out that, despite explicit commitment, such a view implies an implicit rejection of the mechanistic explanation framework widely adopted in traditional cognitive science (TCS), which, in our view, hinders comparability between REC and these approaches. We further argue that such a position is unnecessary and that enactive mechanistic explanation of sociality is both possible and desirable. We examine three distinct objections from REC against mechanistic explanation, which we dub the decomposability, causality and extended cognition worries. In each case we show that these complaints can be alleviated by either appreciation of the full scope of the mechanistic account or adjustments on both mechanistic and REC sides of the debate.

Keywords Enactivism · Mechanistic explanation · Social cognition

✉ Ekaterina Abramova
e.abramova@ftr.ru.nl

Marc Slors
m.slors@ftr.ru.nl

[1] Faculty of Philosophy, Theology and Religious Studies, Radboud University Nijmegen, Erasmusplein 1, 6525 HT Nijmegen, The Netherlands

1 Introduction

The study of social cognition and social interaction is typically aimed at understanding how people manage to deal with other people; how they perceive, understand and predict their behavior, coordinate with them, plan and execute joint actions. In addition to a variety of situations that can be studied and specific theories that have been proposed to account for social phenomena, there is also a variety in how such phenomena should be conceived more broadly and what amounts to a good explanation of any given case. Although theoretical and methodological diversity is generally positive and desirable in science, when the differences run so deep as to make competing explanations of a phenomenon incommensurable, progress can be hindered. This is currently happening in the study of social cognition, with proponents of what we will call 'traditional cognitive science' (TCS) on the one hand and a radical embodied cognition alternative (REC) on the other. These approaches adopt wildly different explanatory frameworks and continually talk past each other. In this paper we aim to illustrate the differences that lead to this impasse, examine its sources and propose a way toward a reconciliation.

One way in which competing views *within* TCS are made comparable is by distinguishing between what a given cognitive capacity allows an organism to do and how it allows the organism to do it. This distinction is inspired by Marr's (1982) proposal on how cognitive systems are to be explained, namely that a complete cognitive theory should specify the system's operation on three levels: computational (what is the system computing? what task is it performing?), algorithmic (what algorithms and representations are used to perform the task?) and implementational (where are the required computations and representations found in the physical hardware of the system?). Since the vocabulary of 'computation' and 'representation' are unnecessarily restrictive, some cognitive scientists have since proposed that Marr's distinction is more fruitfully employed if we focus on the questions that are associated with his levels. For instance, Geurts and Rubio-Fernández (2015) distinguished between W- and H-level, where 'W' stands for *what* the system is doing and *why* and 'H' for *how* it is doing it. While an answer to a what-question is typically a description of a system's behavior that is a target of one's explanation, an answer to a how-question is a proposal as to what states, operations, transformations, components etc. are involved in producing such a behavior.

Although there is a number of ways in which to relate Marr's scheme to explanations that are given in terms of *mechanisms* (e.g., Piccinini and Craver 2011; Bechtel and Shagrir 2015; Zednik 2017), the one we will use here is to say that the what-level serves as a description of an explanandum phenomenon while the how-level is a proposed mechanism that is aimed at explaining that phenomenon. Viewed this way, theoretical development *within* TCS can be said to proceed by competing refinements of the descriptions of target phenomena – for instance, by sketching competing functional analyzes of some capacity that might be exhibited by the organism (Craver 2006) – or by proposing competing mechanisms that actually realize these phenomena. A particular TCS account of, say, memory, vision or social cognition is typically a combination of the two levels and new evidence is collected and interpreted as supporting one of such accounts (or rather, as refuting the alternatives). Of course, it

might also turn out that what seems like competing positions are in fact compatible because the mechanisms that implement them are (or can be) co-instantiated.[1]

We admit that many enactivists might wish to claim that they are not interested in being comparable with TCS, especially if it requires them to adhere to Marr-inspired levels of analysis. They might say that they want to overhaul the TCS framework altogether and propose a new standard for what a complete explanation of a piece of cognition should look like and for how one should connect observed adaptive behaviors with their biological (or artificial) counterparts. However, we feel that in the interest of general discussion among everybody interested in understanding cognition, we should at least investigate whether there are principled reasons to resist the possibility of a REC-y how-level that could be directly compared to TCS alternatives.[2]

Thus, we claim that it is in the interest of enactive accounts of human sociality to see whether and to what extent it can incorporate how-level mechanisms. It is one thing to have a good motivation for wanting REC accounts of sociality to allow for a mechanistic how-level counterpart, though, but it is quite another to show that a mechanistic how-level for REC accounts of social cognition is feasible. This is what we aim to do in this paper (Section 4). There are three worries that motivate enactivists to resist mechanisms. First, mechanisms are conceived of as being fully decomposable and hence reducible to components whereas enactivists reject such reductionism. Secondly, mechanisms are thought not to allow for the kind of inter-level causal interaction that plays a crucial part in enactivism. Thirdly, mechanistic explanation is often associated with brain-centered, non-extended cognition. We shall argue that none of these consideration provides an insurmountable obstacle to a mechanistic how-level account of enactive social cognition. But before that, we will first properly introduce REC accounts of sociality in general (Section 2.1) and enactivist accounts more specifically (Section 2.2), as well as the idea of mechanistic explanation (Section 3).

2 Stage setting

2.1 Varieties of embodied sociality

In order to introduce the various forms of embodied social cognition, it is useful to first sketch the (still popular) 'traditional cognitive science' (TCS) approach to

[1] This strategy of comparing theories of cognition in TCS may but need not presuppose that all explanations are in principle reducible to mechanistic explanations. It all depends on how we view the relation between what- and how-level explanations. If the relation is regarded as strict implementation, then this opens up the way to mechanistic reduction. But when the relation is viewed in terms of idealization or interpretation (in which the what-level approximates the how-level; see e.g. Dennett (1987), such reduction is not implied. We shall set the issue of explanatory reduction aside.

[2] If nothing else, this paper should be taken as an invitation to the enactivists to provide explicit arguments against how-level explanations as well as an explicit commitment to what model of explanation enactivism puts forth as an alternative.

social cognition as a contrast class. On the what-level of description, TCS assumes that humans can interact with others successfully only if they are able to see them as beings with hidden mental states, which they can infer from observable behavior. Such inferences either involve a so-called 'theory of mind' or simulation routines. The results of these inferences are thought to be plugged into the planning of the agent's own actions. There is a stunning variety of TCS views on social cognition, but they all share a general commitment that cognition is about processing information – perceived behavior of the other – by the brain.

This coarse-grained what-level characterization of social cognition suggests the rough outlines of the how-level mechanisms underlying social cognition, according to TCS. Figure 1 schematically depicts the way in which social interaction is facilitated by the interacting agents' representation of each other's motivational states, broadly conceived (including emotional, intentional and epistemic states). This figure is still to be interpreted as a what-level – the cogs suggest a first step in functionally analyzing the phenomenon so as to make it susceptible to how-level mechanistic explanation – see Section 3. The non-blurry cogs represent the motivational states of the agents and the blurry cogs represent their representations of the other agent's motivations. Traditionally, motivational states are thought of as (folk-psychological) mental states. This is the case in so-called theory theories (see e.g., Gopnik and Meltzoff 1997; Stich and Ravenscroft 1992) and most versions of the simulation theory (Goldman 1989, 2006; Gordon 1986, 1996; accepts this too but (Gordon 2008) tends to be sympathetic to resonance-based simulation theories too – see below). Thus, in Fig. 1 the cogs represent mental states and representations of mental states (e.g. beliefs and desires). This means that the actual neural mechanisms that constitute an agent's motivations and the neural mechanisms that constitute the other agent's representations of these motivations differ quite considerably (this is conveyed by the blurriness of the representation cogs). The unfolding of interaction between agents is explained, by all TCS theories, in terms of the operations of this internal

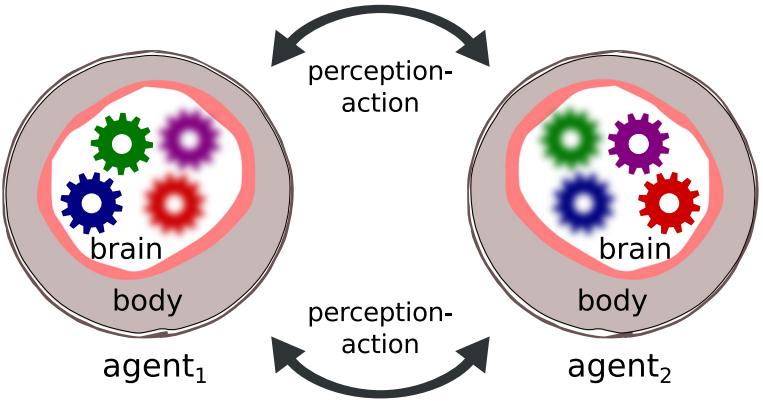

Fig. 1 Traditional cognitive science. Differently colored cogs depict the states of the agents' minds/brains. All the states of the agents are located inside their minds/brains and not their bodies. The clear cogs are the agent's own states while blurry cogs are representations of the states of the other agent

representational machinery, allowing for the process to even proceed offline, without continuous access to information about the other person.

Embodied approaches to cognition (EC) have emerged as an alternative to traditional cognitive science in the early 90s and have since been gaining ground while also diversifying in their particular commitments. In terms of intensity of their opposition to TCS, one can distinguish two families of EC: call them weak EC and radical EC. A weak EC view is most generally a plea to acknowledge that the states of the body and the environment can influence cognition and that lower sensorimotor knowledge plays a role in higher cognition like language and reasoning (Dijk et al. 2008). Such a view emphasizes real-time interaction with other people and perceptual information available in such settings. However, in many theories within weak EC, the brain still plays a central role, as shown in Fig. 2. That is, the body and the environment of an observed agent matter only insofar as they are represented by the brain of the observer. Certain varieties of simulation accounts of mindreading fit into this framework. For example, the appeal to mirror neurons as mediating social understanding still requires that one person replicates the state of the other person in their head. These states need not be mental states, like in Fig. 1. Some motivational states are mental, others are motor processes and ensuing bodily movements.

In the figure, bodily movements and motor processes are depicted as one cog in the body. The cog in the observer representing this bodily cog of the person observed is less blurry than the cog representing their mental state for two reasons. On the one hand, bodily movements are easier to access than mental states. On the other hand, so-called 'motor-resonance' (Gallese et al. 2004) or 'unmediated resonance theories' (see Goldman and Sripada 2005) claim that neural processes driving the behavior of an observed agent are partly replicated in the observers' brain (the replication is only partial and it is taken offline, so that it does not directly cause the observer's behavior).

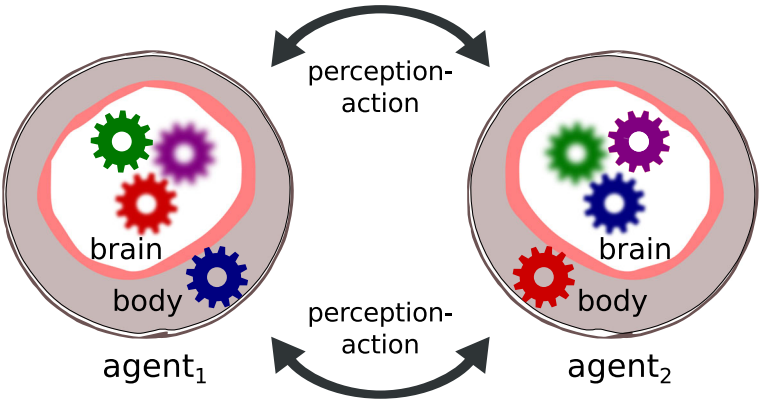

Fig. 2 Weak embodied cognition. One agent has her own bodily state (blue cog for agent$_1$) and her own mind/brain state (green cog for agent$_1$) but is also representing the bodily state of the other agent (somewhat blurry red cog in agent$_1$ which corresponds to the red cog in the body of the agent$_2$) and their mind/brain state (very blurry purple cog in agent$_1$which corresponds to the purple cog in the mind/brain of agent$_2$)

A Radical Embodied Cognition (REC) approach is the view that the body and the environment are actually *part of cognition* and as a result there is no need to have internal representations of the environment, other people or their perspective on the world in order to coordinate with them successfully (Wilson and Golonka 2013). However, this basic claim has been further developed in two different ways. Some REC-ers focus on exploring the emphasis on real-time interaction with other people instead of detached theorizing about or simulating them. This often goes together with a view that social cognition can be realized through "direct social perception" of other people's mental states. That is, for example, perceiving somebody's emotional expression elicits your own response directly, without having to simulate or interpret it first (Gallagher 2008). Since this view assumes that the body is literally part of the cognitive system, while the agents have perceptual access to each other's bodies they thereby access each other's minds (note the lack of blurriness in Fig. 3). There has not been an explicit commitment on the part of REC-ers of this type as to what explanatory framework does justice to their views. Given their general focus on perception-action systems, however, one could surmise that the main questions that guide research of this type have to do with the perceptual input available to the agent in a social situation and the way this input directly guides the agent's response (expressed in thicker perception-action arrows in the figure).

Other REC-ers focus on a different point, namely that the particular dynamics of social interaction as such play a crucial role in explaining social cognition. This is because "becoming a temporary unit of social action with another person also involves creation of a new perception-action system with new capabilities" (Marsh et al. 2009, p. 1219). Therefore, the correct level of analysis in the study of social cognition is the social unit, rather than individuals that comprise it and their internal cogs. Instead of examining properties of individual independent cognizers (be it their brains or bodies), we are to investigate the social interconnectivity that emerges as a result of the interaction and constrains individual-level behavior from the level of a

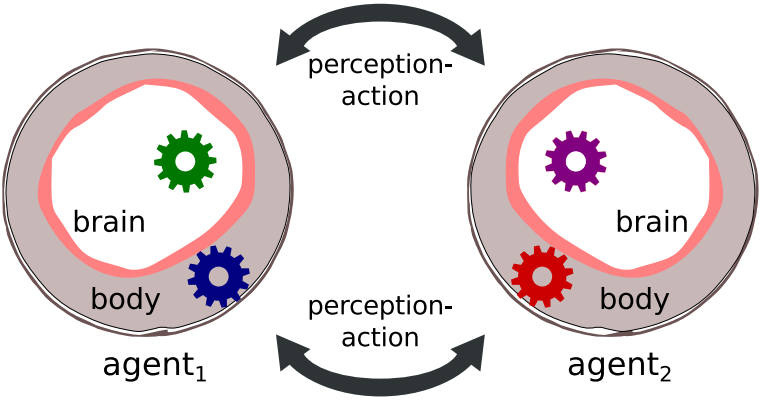

Fig. 3 Radical EC, Direct Perception View. Both agents have only their own states of their bodies and minds/brains and no representations of the states of the other agent. Perception-action arrows are thick depicting the importance of this process in the interaction

new overarching structure. That is, although it is acknowledged that there is something to be said about the individual brains and bodies (the blurry blobs in Fig. 4), an explanatory search for cognitive mechanisms is rejected, typically in favor of dynamical analyzes of social interaction on a higher level.

From the survey of these positions on social cognition it is clear that TCS and REC — especially its second variety — are not just different theories that purport to explain the same phenomenon. Rather, they adopt a different what-level understanding of a target phenomenon (cognition in general and more specifically social cognition or social interaction) and a different view on what it even means to provide an explanation. We believe this unbridgeable gap is not a necessary state of affairs and that even the more extreme version of REC is in fact compatible with a mechanistic how-level focus of traditional cognitive science. Before we move on to this argument, however, a short presentation of the supra-personal approach and the general issue of explanation in cognitive science is in order.

2.2 Enacted sociality

The supra-personal view on social cognition is advocated by two main sub-groups within REC: advocates of complex systems approach to cognition and enactivists. These two strands of REC share many theoretical and methodological commitments. However, enactivism is a strand that is more specifically about cognition[3] and therefore we will will focus on enactivist take on sociality in this paper.

Enactivism stems from the early work in philosophy of biology of Maturana and Varela (1980b) and was popularized as an alternative to traditional cognitive science by Varela and Thompson (1991). It shares theoretical commitments with complex systems theory, phenomenology and the Buddhist tradition in, on the one hand, grounding cognition on the organizational principles of living systems while at the same time giving a prominent role to the investigation of human experience. Three main principles adopted by enactivism are (1) challenging the dichotomy between internal components of the system and its external conditions, instead stressing the interaction between the two, (2) emphasizing properties of higher (emergent) levels of organization while precluding the possibility of reduction from higher to lower levels and (3) viewing the organism as an active autonomous entity that is geared toward adaptively maintaining itself in the environment.

Recent years have witnessed a development of a specifically enactivist take on sociality, in situations in which two autonomous agents interact with each other. De Jaegher et Di Paolo (2007, p. 493) provide the following definition of a *social interaction*:

> 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

[3]That is to say, complexity science is a field of study dedicated to a variety of physical and biological phenomena, not just cognition.

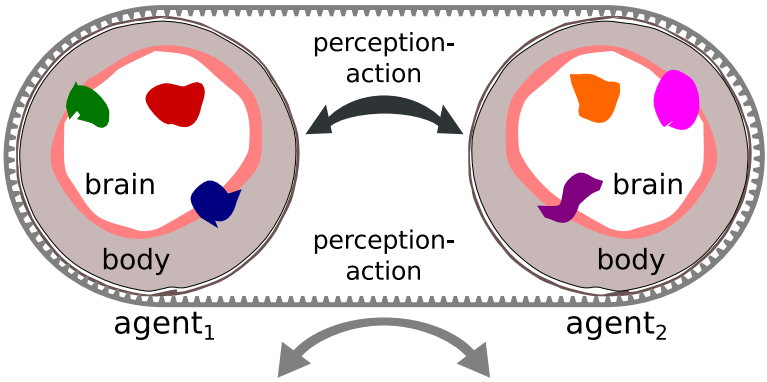

Fig. 4 Radical EC, Supra-Personal View. The agents have some states in their minds/brains and bodies. They are not easily identified nor ascribed to either the brain or the body. There are no representations of each other's states. In addition to perception-action links between the agents, there is an overarching coupling connection (the toothed belt) which gives the whole unity an emerging perception-action opening to the environment (in gray)

> 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).

"Coupling" in this definition is a technical term drawn from the science of complexity. It means that the states of one agent are a function of the states of another agent and vice versa. Although one could say one agent representing the other is also a function of this type, when used by REC-ers coupling is a more basic phenomenon that precludes a need for representation. For example, two oscillating clocks when hanged on the same wall synchronize over time because they are coupled. Their link is not representation-based, it is more direct. The same type of coupling is implied in social interaction.

Thus, an enactivist view on cognition tends to give a different what-level description of social-cognitive phenomena. While one could begin an investigation at the how-level into what exactly becomes coupled between agents, what enactivism emphasizes instead is that the coupling can become self-sustaining and influence the interactants from the higher level. That is, the interaction itself can be viewed as autonomous. In an oft-cited example, imagine two people trying to walk past each other in a narrow corridor and getting trapped in mirroring each other movements. In such a scenario both individuals are autonomous and both have individual intentions to end the interaction and keep walking. However, the nature of the emerging social dynamics is such that their individual intentions get over-ridden and interaction continues.

One of the most distinctive empirical paradigms that exemplifies this idea of autonomous interaction is what is known as a perceptual crossing (PC) study (Auvray et al. 2009; Auvray and Rohde 2012) shown in Fig. 5. In this set-up, a pair of blindfolded participants (call them A and B) are each equipped with a computer mouse and a tactile feedback pad. The mouses correspond, for each participant, with an avatar

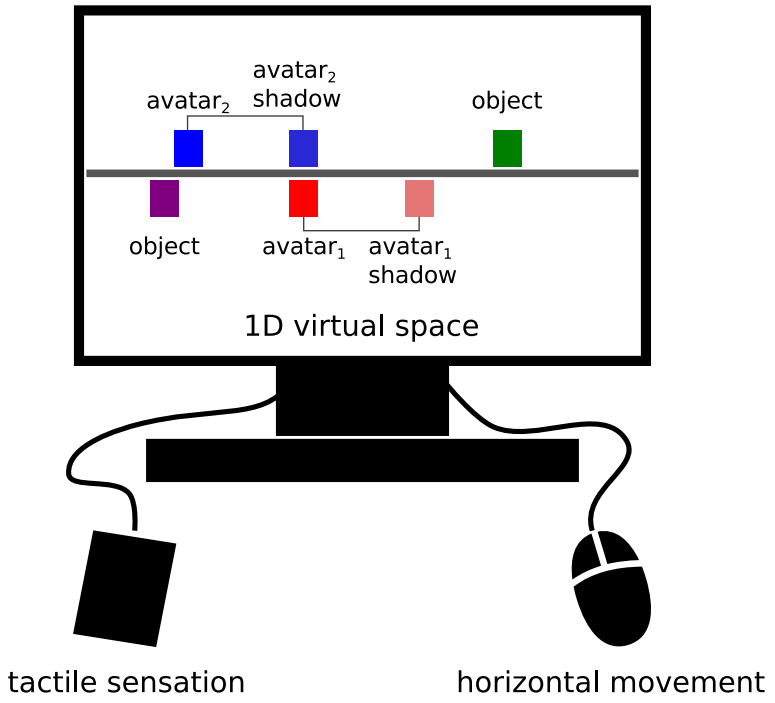

Fig. 5 Perceptual crossing experiment. Each participant in the experiment has an ability to move and receive tactile sensations. They objects shown here on the computer screen represent the state of the 1D environment available to them but cannot be visually perceived by the participants themselves. The objective of each participant is to click the button when they think they are interacting with the other participant's avatar

on a computer screen that can move along a horizontal line. There can be objects at various points on that line. Avatars can 'interact' with objects and 'interact' with each other when they are at the exact same location (since the avatars and objects have a certain width, so does a location). Whenever an avatar interacts with an object or another avatar, the (blindfolded) participant that operates this avatar will feel this via a vibration in the tactile feedback pad. Avatars and objects will elicit the same tactile feedback, but a participant might be able to tell whether she (her avatar) interacts with an object or another avatar by being sensitive to the different patterns of interaction induced by objects compared to avatars. That is, since avatars move and objects remain in a stable location, the latter will elicit consistent vibration as participant moves along its (1D) "shape". On the other hand, if participant remains still but vibration feedback changes, it most likely means that an avatar of the other participant has been encountered. Metaphorically, we might say that avatars are 'alive' because they move and feel, while objects are 'dead' because they do neither. Now there is a complicating factor: both avatars have 'shadows'. A shadow is located a small distance from the avatar and it moves exactly parallel to it (i.e. the distance between avatar and shadow remains the same). Shadows move, then, but they do *not* feel: when an avatar of one participant bumps into a shadow of another participant,

the participant operating the avatar will feel a vibration on her pad, while the participant operating the shadow will feel nothing. In that sense shadows are as dead as objects. Thus: avatar-object and avatar-shadow interactions are felt by the avatar-owner while avatar-avatar interactions are felt by both avatar-owners. The assignment that both participants get is to figure out when they interact with an avatar rather than a shadow or an object. Once they do, they are asked to click a button.

Results show that participants were in some sense successful at this discrimination. Both participants adopt a scanning strategy, moving back and forth when feeling an interaction. Using such a strategy, the difference between a static item (object) and a mobile item (other agent or shadow) can be made. But the distinction between encountering an avatar and its shadow is much harder: An avatar that is 'being scanned' by another avatar will feel continuous stimulation even when she does not move herself. But the same sensation can be the result of a shadow moving back and forth over the avatar due to the fact that, for instance, the shadow's-avatar is scanning a static object a bit further away.

What makes the experiment so interesting are the paradoxical results. On the one hand, in the majority of cases in which participants indicated they were interacting with another avatar, they were in fact correct (i.e., the absolute number of clicks was in fact preceded by an interaction with an avatar). On the other hand, participants were nearly as likely to guess they were interacting with an avatar when they were indeed interacting with an avatar as when they were interacting with their shadow (i.e. the number of clicks relative to the proportion of interactions with avatar compared to shadow did not differ). This paradox is easily explained: avatar-avatar interactions were more frequent than avatar-shadow interactions. The situation of 'sensing the other' while 'being sensed' turns out to be more stable than sensing an insensitive shadow. The task is solved globally, then, even if participants are not conscious of this effect[4].

This result is important. A frequent assumption in TCS explanations of social cognition is that recognizing another as a minded being is accomplished by some special cognitive capacity, often referred to as 'mindreading' or agency detection, and that it is a precondition for interacting successfully. Applied to the PC experiment this would be like assuming that the participants have a capacity to infer when the other participant felt their interaction too and when not. People do not have such telepathic capacities, so from a TCS point of view they should not be able to tell whether the item their avatar interacts with is another avatar or a shadow. TCS theorists would therefore most likely emphasize that "the relative recognition rate, i.e., the ratio of clicks per type of object divided by stimulations per type of object, does not differ between the mobile object and the interaction partner" (Auvray and Rohde 2012, p. 2)

[4]The divergence in results is often difficult to grasp. To explain this in another way: imagine a bag of black and white marbles and your task is to take out a marble and guess its color but the only answer you can give is 'white'. This is analogous to the fact that participants had to only respond when they thought they were interacting with the other's avatar, not guess differentially what entity they are interacting with. You take out marbles and say 'white' and it turns out you are correct 2/3 of the time. However, it also so happens that 2/3 of the marbles in the bag are actually white. Thus, you are correct majority of the times but mostly due to the distribution of your chances to be so.

and so participants failed at the task. Enactivists and other REC theorists, by contrast, emphasize that "they clicked significantly more often when meeting the partner's avatar" than in other cases and so "were able to perform this task well" (p. 2). That is, from a supra-personal perspective the task is solved by the fact that avatar-avatar interactions are more stable, more self-perpetuating, than avatar-shadow interactions. For this results in the fact that overall significantly more clicks were correct than false. From a REC perspective, the social interaction itself and its particular dynamics constitute a solution to a task of agency detection. Therefore, in the oft-repeated claim by enactivists, *social interaction constitutes social cognition*.

The difference between the approaches to this experiment is a difference in what-level description of social-cognitive tasks. According to TCS, agency detection is a matter of being able to fathom another person's mind; according to enactivists, agency detection is a matter of being engaged in self-perpetuating, stable interactions (in the case of this experiment). On the TCS what-level characterization, a how-level explanation should focus on individual social-cognitive mechanisms. On the enactivist what-level description, the issue of how-level explanations is much more complex. This is what we will turn to in the following sections.

3 Mechanisms and how-level explanation

What enactivists claim with respect to explanations of sociality is that in many instances in which TCS invokes mindreading, interaction supra-personal processes such as the stability of avatar-avatar interactions in the PC experiment will do too. In fact, the PC experiment shows that this description is to be preferred since it is better at capturing the pattern of results. Thus, the experiment suggests that we should at least take seriously the idea that the supra-personal interaction-based social cognition promoted by enactivists plays a more prominent role than TCS supposes. And yet, enactivist social cognition does not play a significant role in mainstream social cognition research. As we will explain in this section, the reason for this is that when it comes to explanations, enactivists content themselves with precise, prediction-supporting descriptions of the what-level. TCS explanations, by contrast, allow for mechanistic how-level explanations. In the absence of a similar how-level counterpart of enactivist characterizations of social cognition, mainstream researchers see the REC what-level characterizations as incomplete and incomparable to their own. To state this differently, when comparing options, a what-level account that does allow for a how-level explanation tends to be preferred over a what-level description that does not come with such a how-level explanation (and does not even allow for one). Let us elaborate briefly on this.

Phenomena such as the apparent stability of avatar-avatar interactions can be described and modeled in precise ways. Enactivists employ sophisticated means such as artificial agent simulations and dynamical systems theory for this. For instance, Di Paolo et al. (2008) and Froese and Di paolo (2010) implemented the PC study in agents controlled by a neural network and trained to perform the task in an artificial evolution. They confirmed that it can be solved using very basic resources (the agents are very simple and have no mindreading module) and that the best explanation lies

indeed in the stability of the interaction and not in the individual perceptuo-motor capacities of the agents. In fact, the latter study (Froese and Di paolo 2010) showed that the influence of the interaction dynamics is robust even in a situation in which individual capacity cannot contribute to solving the task (because the agents are wired in such a way as to receive their co-actor's perceptual input) and when it goes against individual intentions (the agents are required to stay with the co-actor's shadow but still end up trapped in the mutual interaction).

What is most interesting about these modeling studies, for the purpose of our paper, is the type of explanation that they put forward. There are three elements that enactivists emphasize. First, they claim that both the PC experiment itself and the modeling examples "point to the dynamics of the interaction process as the explanation of coordinated crossing between subjects and not to an individual sensitivity to social contingency" (Di Paolo et al. 2008). That is, the what-level consists in interacting individuals and patterns that emerge between them and not in their individual behaviors.

Second, the modeling studies emphasize that simulated agents are not to be thought of as models of human behavior or psychology in the experiment. They are rather a conceptual tool to explore the constraints of the task and probe its possible solutions allowing the researcher to challenge the preconceptions about how the behavior must be generated. In fact, there seems to be no simple way to relate the structures and processes that generate PC solution in the simulated agents versus the ones operational in humans.

Third, Froese and Di paolo (2010) carry out a detailed dynamical analysis of the agents' behavior and compare it to assuming that agents are controlled by a circuit that calculates a discrimination decision based on something like the length of the simulation from the objects it encounters. This analysis shows that such a circuit would not be able to explain the results and that instead, it is more useful to consider the dynamics of the agents' internal states, their movement in the environment and responsiveness of the other agent. The explanation is provided in terms of dynamical landscape, attractors, perturbation and hysteriesis. Such notions places this account in the species of dynamical explanation that is traditionally viewed as opposed to a mechanistic explanation (although it does not need to be, see e.g. Zednik 2011) and that raises a typical complaint that what is being offered is merely a finer-grained description of the what-level and not a how-level at all.

TCS descriptions of the what-level of social cognition, usually in terms of some variant of mindreading, do not easily allow for mathematical modeling of this type. They do allow for a how-level counterpart, though, when they are combined with what is known as mechanistic approaches to cognition. In a mechanistic framework, one does not explain a given phenomenon in terms of dynamic patterns that generalize over large terrains, but in terms of the structural nature of (causal) interactions between particular elements that comprise the phenomenon. This is known as identifying a mechanism, where:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena (Bechtel and Abrahamsen 2005, p.423).

Mechanisms implement the processes that are described at the what-level. They are the paradigmatic how-level counterparts of the what-level. A mechanistic explanation proceeds by identifying the phenomenon of interest and then trying to understand how it results from an orchestrated operation of lower-level components and their temporal and spatial organization (Bechtel and Abrahamsen 2005; Bechtel 2010). For example, spatial memory as a cognitive capacity is explained by the structure and functioning of the hippocampus and its connection to other brain regions, as well as the mechanism of long-term potentiation (a form of synaptic plasticity) on yet lower level (Craver 2002).

The mechanistic explanatory strategy has been described by Bechtel (2009b) as the activity of looking down, around and up, i.e., respectively, decomposing a phenomenon into its working parts and operations, establishing their organization (understanding how the parts relate to each other) and situating the mechanism in a larger context. The latter might be a mechanism on a higher level or the environment.

Enactivists do not employ mechanistic explanations. In part this is because mechanism is associated with functionalism and reductionism, to which enactivists are fundamentally opposed (Raimondi 2014). More specifically, there are three main reasons why enactivists avoid mechanistic how-level explanations. First they assume that mechanists necessarily claim that cognitive phenomena can be reduced to a composite of their parts. This would preclude the idea that e.g. the supra-personal level of description has an autonomous explanatory role to play. Secondly, enactivists assume that there is inter-level causal influence – for example, the supra-personal level phenomenon of the stability of avatar-avatar interactions causes individuals in the PC experiment to behave in certain ways. Such inter-level causality is usually denied by mechanists. Thirdly, mechanistic explanations in cognitive science are as yet applied solely at the individual level, while enactivists assume a prominent role for the supra-personal level in explaining social cognition.

Avoiding mechanisms altogether, however, puts enactivist positions on social interaction in second place relative to TCS approaches in the eyes of many, for the simple reason that TCS approaches can and do readily help themselves to various models of the cognitive mechanisms that underlie the individual information-based processes they postulate. The (implicit) rejection of mechanism is thus an obstacle to the wider acceptance and further development of the position. What we want to suggest here is that mechanistic explanation is not only compatible with enactivism but also preferable.

Our proposed version of mechanistic enactive explanation, depicted in Fig. 6, combines both versions of REC outlined above. The what-level describes kinds of social interactions in which individuals participate. The how-level explanation is to be achieved by specifying all the components of the picture that contribute to the realization of such interactions. The components of the cognitive mechanisms (the cogs) are distributed across the brain and the body of both agents embracing the version of REC that emphasizes direct perception (Fig. 3). They are also dynamically coupled (the toothed belt), respecting the enactivist rejection of the internal-external dichotomy. In contrast to the pure enactivist view (Fig. 4), what the current picture stresses is that individual brain-body cogs are also part of the picture and are required for a complete explanation. Their contribution, however, is diminished with respect

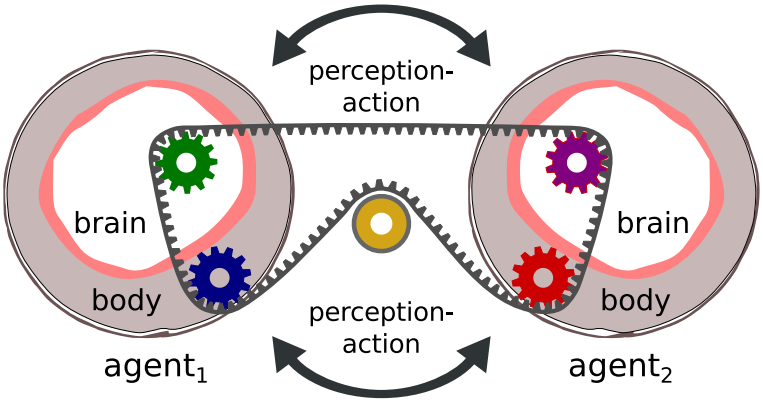

Fig. 6 Enactive mechanistic view

to Figs. 1 or 2, suggesting a need for an alternative account of such internal mechanisms. The fact that the coupling is a constraint on individual mechanisms rather than an additional cog, expresses the idea that interaction consists of interacting individuals yet allowing for emergent effects resulting from the linkage. Furthermore, the picture includes the possibility that the coupling might be affected by contextual factors (the tension pulley), such as the layout of the environment in which interaction unfolds, or some socio-cultural circumstances.

Relating the picture to the PC experiment, the enactivist mechanistic view would imply that an explanation for the supra-personal what-level effect observed therein (the superior stability of avatar-avatar interactions relative to avatar-shadow interactions) is right to focus on the inter-connectivity between the participants. However, what it adds is the need for taking into account the particular environmental constraints present in the task (the one-dimensional environment, the types of objects that can be encountered) and the particular sensory and motor capabilities of the participants (moving the mouse, sensing vibrations). These capabilities are parts of the individuals, of their bodies and brains and an explanation is not complete without specifying their contribution. The various components and processes contained in the task description above are tightly interconnected in concrete causal terms. This does not preclude, however, that the whole set-up, including the two persons and the equipment, can be described as one mechanism relative to which the various parts play their role. It is impossible to describe this mechanism in a paper such as this (if indeed at all, since, for one thing, the relevant brain components may as yet be unknown). But the principles underlying such a description are captured intuitively in the Fig. 6 scheme and are as such, we think, sufficiently distinguishable from established enactivist approaches.

The acceptance of our enactive mechanistic picture hinges on the possibility of combining enactivist with mechanistic explanation. Thus, in the remainder of the paper, we will consider the three main objections raised by REC researchers against mechanistic explanations and show that they rely on a misunderstanding of what such an explanation entails.

4 Enactivist worries about mechanistic explanation

We will examine the three main worries that have been raised so far that conveniently fall into the main categories of questions that guide a mechanist, what Bechtel has called looking down, around and up. We first discuss the decomposability worry that concerns Bechtel's looking down and around, namely the possibility of decomposing higher-level phenomena into lower-level components organized in a certain way, which enactivists consider reductionist. We reply that such a view relies on a misconception of the mechanistic approach.

The second worry we turn to is related to the effect of higher-level emergent levels onto lower level constituents (looking up in the sense of situating the components inside a larger mechanism). Enactivists fear that a mechanistic approach precludes such top-down causality which they think essential to the autonomy (i.e. the non-reducibility) of the social level and the possibility of so-called 'circular causation'. Most mechanists do indeed reject inter-level causation. We argue, however, that a recent proposal to explain how mechanism can allow for inter-level causation is convincing and fully compatible with enactivism.

Finally, the extended cognition worry has to do with looking up in situating the individual in a social and physical context. We contend that the core mechanistic literature does not seem to cover extended and supra-personal mechanisms, as required by enactivism. However, recent developments in the field show that there are ways to create such a possibility while staying true to both mechanistic framework and enactivism.

4.1 The decomposability worry

The main worry enactivists and other REC-ers seem to have with the mechanistic approach is that it allegedly views cognitive systems as decomposable or near-decomposable while in reality they are non-decomposable. For example, Lamb and Chemero (2014) argue that according to the mechanists, producing an explanation requires (1) "decomposition [that] involves developing a model of a system's behavior by identifying discrete component parts and their linear, or weakly non-linear, interactions" and (2) "localization [that] involves mapping those discrete components and interactions onto features of a physical system" (pp. 809-810). What is often added to this charge is that such an explanatory strategy views cognitive systems as *component-dominant*, i.e. the behavior of the whole is a simple additive result of the behavior of its components, whose properties and functions are rigid and pre-determined (Favela 2015). Therefore, a single component can be analyzed in isolation as responsible for some particular capacity of the system. Applied to the brain, for example, it would mean that we can identify and localize particular brain modules responsible for particular cognitive tasks like vision, processing information about other agents, reading written text and so on. Taking out that part of the brain or disrupting it would mean that the whole system loses that particular capacity.

In an opposition to this view on the brain and cognition, REC-ers argue that in fact living cognitive systems are non-decomposable into components and *interaction-dominant*. That is, the behavior of the whole is more than a simple sum of the parts

because interactions between parts are mostly non-linear, the behavior of each part dynamically depends on all other parts of the system and it is not possible to assign any specific task to any component. Therefore, interactions between components are more important than the components themselves (Richardson and Chemero 2014). Viewed through that lens, it is not possible to analyze the brain and cognition into separable modules. Removing a part of the brain will obviously lead to some loss of function. However, this is not because the brain lost a particular component that realizes a particular property but because the brain is then a different whole operating differently (Maturana 1980a). One should note that we have not claimed that enactivists or interaction-dominant explanations deny the existence of components altogether. What they do is deny identification of components and their contribution to the realization of the phenomenon as an important part of the explanation of this phenomenon. Differently put, the contribution of a component to the overall behavior of a system is not tractable or identifiable in terms of the taxonomy that is appropriate to describe the overall behavior of the system (much like this is the case with, say, the contribution of genes to the complex behavior of animals).

As can be seen from the comparison between a component-dominant and interaction-dominant view on the cognitive system, they are two extremes of a continuum of positions that might be held by supporters of REC. Rejecting a component-based explanation could mean rejecting explanations that (a) take into account parts only but not their configuration, (b) take into account only parts that interact linearly or (c) statically but not dynamically, (d) take into account all sorts of interactions in addition to parts but not the modulating effect of the environment.[5] It is unclear at present which of these options enactivists subscribe to. However, we can examine where the mechanistic approach places itself on this continuum.

If neural and cognitive systems are indeed non-decomposable and the mechanistic framework can only be applied to decomposable systems, then obviously enactivists cannot make use of it. However, these arguments betray a misunderstanding of the mechanistic framework and explicit dismissal of the new developments in this field.

First of all, mechanists explicitly argue against mere aggregation of components and place heavy emphasis on their organization (Wimsatt 1997). It is because the way parts are organized in space and time that they *together* can exhibit behavior that they cannot exhibit on their own. It is because the parts are on a lower level than the whole they comprise that they cannot have the same properties (the properties of the hydrogen and the oxygen atoms are clearly not the same as the properties of water molecules).

Second, there is no reason to suppose that only linear and sequential modes of organization are allowed in mechanisms. Especially when dealing with biological mechanisms, non-linear and cyclic modes are ever-present. Such a focus on biology has led mechanists to stress the necessity for *dynamic mechanistic explanation* because in a system organized non-linearly "the operations performed by parts of the mechanism

[5]We thank a reviewer of a previous version of this paper for pointing out this ambiguity and a range of possibilities that need to be considered.

vary dynamically, depending on activity elsewhere in the mechanism" (Bechtel 2011, p. 551). Therefore, an explanation has to include not just a static diagram of components and their organization but also a description of how the functioning of these parts is orchestrated in time.

Finally, Bechtel (1997), in response to REC challenges has argued that cognitive systems are likely to lie on a continuum between the extremes of non-decomposable and fully-decomposable. They are, instead, *integrated* systems. In such systems, it is still possible to identify components. However, their functions are not necessarily predetermined and fixed. Rather, their contribution to the operation of the whole might dynamically depend on other parts of the system, the larger context and be variable in time. It does not mean that when studying a mechanism for a particular phenomenon it is impossible to identify these contributions.

In reply to such arguments, Lamb and Chemero (2014) state that

> If a neo-mechanist wishes to discard the condition of decomposability, then she does so at the cost of discarding the feature of neo-mechanistic explanations that makes them distinct from more general accounts of naturalistic explanation (p. 813).

We believe this is incorrect. What is distinctive about neo-mechanistic explanations is not decomposability understood by REC-ers to mean "decomposability into linearly interacting static components". What is distinctive about neo-mechanistic explanation is (1) a concern for capturing different levels of the system and understanding the relations within and between levels and (2) a concern for a particular target phenomenon and the concrete working parts and operations that underlie it. The latter property makes mechanistic explanation different from different types of explanation, such as, for instance, dynamic explanation (which seems to be what Lamb and Chemero mean with 'naturalistic explanation') in which generality and an ability to subsume a variety of phenomena under a particular regularity is seen as a virtue. The former property distinguishes mechanistic explanation from a general REC view on inter-level relations that take us into the discussion of inter-level causality, to which we now turn.

4.2 The causality worry

Connected with the decomposability worry is another misgiving of enactivists about mechanisms, which we will call the 'causality worry'. An important part of the enactivist framework is the so-called 'circular causality' that is allegedly operative in many enactive systems. The idea here is that the elements or components that make up a system 'cause' the emergence of properties at a higher level of aggregation that cannot be reduced to the component parts and their interactions. These emergent properties, in turn 'cause' specific effects at the component level, by 'enslaving' components and their properties, as it is called. The PC experiment is a case in point: the overall dynamics of the experimental set-up, including participants, involves the more frequent occurrence of avatar-avatar interactions. This is caused by the actions of individuals, but the overall dynamics of the whole system causes individuals to move their mouses such as to contribute to this effect. Although mechanistic

explanations are keen on levels as well as on causation, causation between levels has not traditionally been part of the mechanistic picture (Bechtel 2008; Craver 2007a, b).

Mechanists typically think of causation as an intra-level phenomenon. Inter-level relations are relations of constitution, according to them, and it would be wrong to put these on a par with causal relations. The problem is that on a mechanistic account higher-level phenomena – system S's Ψ-ing, say – are constituted by the causal interactions of components of a given mechanism – such as component C's Φ-ing. This means that C is a part of S and that C's Φ-ing is part of S's Ψ-ing. Thus, top-down causation in a mechanistic framework would seem to involve causal interactions between a whole and its parts. This is problematic because according to many, if C and S are related as part and whole, they cannot be related as cause and effect. Causes and effects are thought to be (i) wholly distinct, (ii) temporally asymmetric (causes precede effects) and (iii) unidirectionally dependent (effects depend on causes, but not *vice versa*). However, wholes and parts are (i) not wholly distinct, (ii) temporally coincidental, and (iii) dependent in a direction (wholes are constituted by parts, not vice versa), that is incompatible with causal dependency in top-down causation (where parts should depend on wholes). For reasons such as these, mechanists reject the idea that the constituent relations between levels leave room for causal relations.

The notion of constitution at play here is synchronic. Or better: it is a notion in which the diachronic nature of processes – whether at the component level or the system level – does not play an explicit role. It is for this reason that Kirchhoff (2015) has argued that the notion of constitution as employed by REC is radically different from the way analytic philosophers, including mechanists, use that notion. Constitution on REC accounts is essentially and fundamentally diachronic; it is the dynamic unfolding of interconnected lower-level processes that constitutes events at a higher level. The notion of constitution that Kirchhoff uses as a contrast class for this dynamic, diachronic constitution, though, is taken from the kind of analytical metaphysics that is not concerned with cognition or processes in the first place: Gibbard's (1975) example of a piece of marble that constituted Michelangelo's David is used as the main model. Though this model has been used to argue that persons are constituted by bodies (Rudder-Baker 2000), it should be clear that a constitution relation that is as static as the relation between a piece of marble and a statue cannot used as be a model for the way in which lower-level processes constitute higher-level cognition. Kirchhoff is right when he claims that diachronicity has been disregarded by mechanists. This is not because they think interconnected components of mechanisms are as static as pieces of marble. It is because they have failed to be explicit about the fact that these components are processes too.

This is exactly what Krickel (2017) has done in a recent paper. By doing so she has killed two birds with one stone: not only is her diachronic notion of constitution a plausible diachronic extension of the standard mechanistic picture drawn by e.g. Bechtel and Craver. More importantly for our discussion, she shows how a diachronic notion of mechanistic constitution, in which interconnected lower-level processes together constitute a higher-level process, makes room for inter-level causation. In order to see how, we first need to distinguish between two ways in which a system-level process can be subdivided in parts. *Temporal* parts of such a process are parts

of the overall system-level process – they are time-slices of such a process. If the process is the process of a person dying – to use a sinister but simple example – a temporal part of it may be the moment in which a person looks shocked and brings his hands to his chest. *Spatial* parts are at the component-level; but like temporal parts they can occur during only a part of the overall system-level process. In the example of a dying person, the event of a heart that stops beating would be a case in point. Inter-level causation becomes possible, according to Krickel, because spatial parts of an overall process and temporal parts of such a process are *not related as parts and wholes*. Suppose that the event of diving in ice cold water and the event of ceasing to move are temporal parts of the process of some person's dying, and that the event of a heart that stops beating is a spatial part of that overall process. The heart that stops beating is not related as a part to either the process of diving into the water or the process of stopping to move. If spatial and temporal parts of a single overall process are not related as parts and wholes, then they can be related as causes and effects: they are distinct and temporally related and they can have asymmetrical dependence relations. In our example it would mean that we can say that the diving in ice-cold water caused the heart to stop beating, which in turn caused the person to stop moving. And these causal relations are all constitutive of the overall process of dying.

Krickel's mechanistic notion of inter-level causation fits our model of mechanistic enactivism (where the tooth-belt of global processes causes movements in the component cogs and *vice versa*). It also fits the enactivist diachronic/process view of constitution. In fact, prominent enactivists cite Krickel's position with approval (Gallagher in press; Gallagher stops short of explicitly accepting diachronic mechanism but he certainly does not reject it).

4.3 The extended cognition worry

The third major worry enactivists could have about mechanisms is that they prevent cognition from being understood as extended, i.e. done not by the brain alone but rather by a brain-body-environment system. In the case of social cognition, it is rather an extended brain-body-environment-body-brain system (Froese et al. 2013).

That the worry is justified is illustrated by the following critique by Herschbach (2012). In his article on social cognition sub-titled "A mechanistic *alternative* to enactivism" (emphasis added), he very acutely points out that enactivists have not been very clear on what they mean by constitution in their claim that social interaction constitutes social cognition. Constitution is a part-whole relationship and if the claim is that supra-personal interaction constitutes individual cognition, then it is somehow a category mistake and a confusion of levels of organization.[6] On the other hand, if constitution is aimed at emphasizing the causal links between agents engaged in

[6] 'Levels of organization', such as a level of an organism, organs and cells, is a different distinction than 'levels of explanation' and we should be careful not to confuse the two. In general, one can provide a what- and how-level explanation for any particular level of organization although sometimes it certainly does seem that in explanations of cognition the what-level behavior of full persons is coupled to how-level mechanisms of person parts.

the interaction, then enactivists are committing a well-known coupling-constitution fallacy (Adams and Aizawa 2010). In this fallacy, frequently ascribed to proponents of extended cognition in general, one points out extensive causal coupling between a cognitive agent and some external factors and then concludes that therefore these factors are part of cognition. Such a conclusion is thought to be unwarranted because coupling and constitutive relations are in general not equivalent.

Herschbach proposes that adopting a mechanistic framework can capture everything that enactivists want to say about social interaction without committing the fallacy. He states that the perceptual crossing example would be described by mechanists as

> an autonomous social network composed of two interacting agents [with some emergent properties to be explained by] (a) decomposing the system into its parts – the agents and potentially other environmental objects – and determining how each part behaves, (b) examining how those parts are organized spatially and temporally to constitute the entire social network, and (c) determining how that network interacts with anything external to it (p. 482).

Enactivists probably would not find this description troubling. However, Herschbach moves on to claim that a mechanist would focus on the lower level of the individual agents and how their internal mechanisms are responsible for the scanning behavior observed in the experiment. This behavior is responsive to the kind of sensory input received by the agent (from another avatar vs their shadow). The main point of difference between enactivists and mechanists, according to Herschbach, is that while the former would like to say that such environmental input constitutes social cognition, the latter would say that only the agent-internal mechanism constitutes the phenomenon of interest (the scanning behavior) while the environmental input is merely an external influence on that mechanism. The mechanism succeeds only when situated in the appropriate social context of having contact with another agent. In conclusion, rather than talking of the constitutive role of the interaction, Herschbach suggests that the emerging pattern is to be explained by internal capacities and dynamics of the agents (this is the constitutive part) that are situated in the appropriate social environment that causally interacts with it.

Why would Herschbach say this? Are mechanists necessarily internalists with respect to cognition? They are not. While most mechanists say little about the issue of extended cognition, Zednik (2011), for one, has argued for a possibility of truly extended mechanisms. He argues that dynamical explanations "are well suited for describing extended mechanisms whose components are distributed across brain, body, and the environment" (p. 239). That is, body and its brain on the one hand and the environment on the other can be said to be the two working parts of the mechanism (see also Beer 2003). Following this idea, Rucińska (2016) adds that the said parts can be conceived in a non-representational manner to fit wider enactivist commitments, by focusing on the 'know-how' in the animal's body and affordances as constituents of the environmental side of the equation.

The link between mechanism and internalism that Herschbach assumes is not inherent in mechanism but rests on a *further* assumption, not about the nature of explanation, but about the nature of cognition. Herschbach thinks that only parts that

participate in a self-organized autonomous individual can be truly said to constitute cognition. He follows Bechtel (2009a) in this claim, who in turn argued that it is the autonomous living system that is the proper "locus of control", differentiated from the environment, because it is the living system that needs to maintain itself as a unity in constantly changing external conditions. Thus, even if we were to regard the whole PC set-up as a large mechanistic system, it would simply not be a cognitive system, according to Herschbach.

The obvious thing to note here is that Herschbach replaces the enactive explanandum – the enactivist what-level description – with his own by switching from the phenomenon of interest being social interaction as a whole to the scanning behavior of the individual – which is the standard TCS what-level description. Like Bechtel, enactivists think that organisms are loci of autonomous control. However, they are autonomous in being operationally closed. This, however, applies not just to the biochemical processes of self-maintenance, but also to the closure of the sensorimotor loop of the organism. This loop is closed not *to* the environment but *through* the environment, which is merely an additional step in the loop, not an input or output external to the system (see Villalobos and Ward 2015, for a more detailed argument). The point here is that if the enactivist notion of autonomy did not allow for the role of the environment in the cognitive process, they could not coherently advance extended cognition type of claims.

Herschbach may reject this enactivist notion of environment-involving autonomy. But he cannot do this on the grounds that it precludes a mechanistic explanation. The fight between Herschbach and the enactivists is not at the how-level where mechanism is at home, but at the what-level.

5 Conclusion

Radical enactivist explanations of social cognition have tended to reject a possibility of how-level explanations. This pushes radical enactivists in a passive defense position relative to classical cognitivist explanations of social cognition, since the latter can avail themselves of a more detailed mechanistic type of explanation. In this paper we have argued that this situation is unnecessary, as mechanistic radically enactive explanations of social cognition are possible too. This, we claim, can put enactivist and cognitivist explanations on equal footing, which would make a more balanced comparison possible. It can allow radical enactivism to become more integrated with the rest of cognitive science. And it can allow radical enactivists to focus on the role of individual cognition in processes of social interaction without giving up on the extended nature of social cognition and the possibility of supra-personal explanation.

For this we have discussed the three main alleged objections from radical enactivists against mechanistic explanation, which we have labeled the decomposability worry, the causality worry and the extended cognition worry. With respect to the decomposability worry we have argued that allowing for mechanisms to be decomposable in components need not turn mechanisms into mere aggregates of linearly interconnected components. On the contrary, it can allow for complex, dynamic, non-sequential interactions that result in emergent system-level properties. With respect

to the causality worry, we have argued that while such emergent properties are constituted by the interconnected components of a given mechanism, they can, in turn, be said to exert influence on these components. We have argued that these mutual relations of influence should not be conceived as causal relations to fit the mechanistic framework and need not be conceived as causal relations to capture the enactivist commitments. Finally, with respect to the extended cognition worry, we have argued that, contrary to what is assumed by many cognitivists, mechanistic explanation does not stand in the way of extended cognition.

We believe enactive mechanistic explanation is definitely possible. All it requires is an appreciation of the full scope of the mechanistic framework and its adjustment to fit wider enactivist commitments. A mechanistic reorientation of radical enactivism can be advantageous to enactivism as such, and can put enactivism in a better position in comparison to traditional cognitivist approaches in cognitive science.

References

Adams, F., & Aizawa, K. (2010). Defending the bounds of cognition. *The Extended Mind*, 67–80.

Auvray, M., Lenay, C., Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas in Psychology*, *27*(1), 32–47.

Auvray, M., & Rohde, M. (2012). Perceptual crossing: the simplest online paradigm. *Frontiers in Human Neuroscience*, *6*, 181.

Bechtel, W. (1997). Dynamics and decomposition: are they compatible. In *Proceedings of the Australasian cognitive science society*.

Bechtel, W. (2008). *Mental mechanisms: philosophical perspectives on cognitive neuroscience*. London: Routledge.

Bechtel, W. (2009a). Explanation: mechanism, modularity, and situated cognition. In Robbins, P., & Aydede, M. (Eds.) *The Cambridge handbook of situated cognition* (pp. 155–170). Cambridge: Cambridge University Press.

Bechtel, W. (2009b). Looking down, around, and up: mechanistic explanation in psychology. *Philosophical Psychology*, *22*(5), 543–564.

Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science*, *78*(4), 533–557.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 421–441.

Bechtel, W.R.C. (2010). *Richardson discovering complexity: decomposition and localization as strategies inscientific research*. Cambridge: MIT Press.

Bechtel, W., & Shagrir, O. (2015). The non-redundant contributions of Marr's three levels of analysis for explaining information-processing mechanisms. *Topics in Cognitive Science*, *7*(2), 312–322.

Beer, R. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, *11*(4), 209–243.

Craver, C. (2002). Interlevel experiments and multilevel mechanisms in the neuroscience of memory. *Philosophy of Science*, *69*(3), 83–97.

Craver, C.F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376.

Craver, C. (2007a). Constitutive explanatory relevance. *Journal of Philosophical Research*, *32*, 3–20.

Craver, C.F., & Bechtel, W. (2007b). Top-down causation without top-down causes. *Biology & Philosophy*, *22*(4), 547–563.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507.

Dennett, D. (1987). *The intentional stance*. Cambridge: MIT Press.
Dijk, J., Kerkhofs, R., Rooij, I., van Haselager, P. (2008). Can there be such a thing as embodied embedded cognitive neuroscience? *Theory & Psychology*, *18*(3), 297–316.
Di Paolo, E., Rohde, M., Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas in Psychology*, *26*(2), 278–294.
Favela, J.r., L. H. (2015). *Understanding cognition via complexity science*. Unpublished doctoral dissertation, University of Cincinnati.
Froese, T., & Di paolo, E. (2010). Modelling social interaction as perceptual crossing: an investigation into the dynamics of the interaction process. *Connection Science*, *22*(1), 43–68.
Froese, T., Iizuka, H., Ikegami, T. (2013). From synthetic modeling of social interaction to dynamic theories of brain–body–environment–body–brain systems. *The Behavioral and Brain Sciences*, *36*(04), 420–421.
Gallagher, S. (2008). Direct perception in the intersubjective context - Social cognition, emotion, and self-consciousness. *Consciousness and Cognition*, *17*(2), 535–543.
Gallagher, S. (in press). *New mechanism and the enactivist concept of constitution. The metaphysics of consciousness*. London: Routledge.
Gallese, V., Rizzolatti, G., Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.
Geurts, B., & Rubio-Fernández, P. (2015). Pragmatics and processing. *Ratio*, *28*(4), 446–469.
Gibbard, A. (1975). Contingent identity. *Journal of Philosophical Logic*, *4*(2), 187–222.
Goldman, A. (1989). Interpretation psychologized. *Mind and Language*, *4*, 161–185.
Goldman, A.I., & Sripada, C. (2005). Simulationist models of face-based emotion recognition. *Cognition*, *94*(3), 193–213.
Goldman, A. (2006). *Simulating minds: the philosophy, psychology, and neuroscience of mindreading*. USA: Oxford University Press.
Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts and theories*. Cambridge: MIT Press.
Gordon, R. (1986). Folk-psychology as simulation. *Mind and Language*, *1*, 159–171.
Gordon, R.M. (1996). 'Radical' simulationism. In Carruthers P., & Smith, P.K. (Eds.) *Theories of theories of mind* (pp. 11–21). Cambridge: Cambridge University Press.
Gordon, R.M. (2008). Beyond mindreading. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, *11*(3), 219–222.
Herschbach, M. (2012). On the role of social interaction in social cognition: a mechanistic alternative to enactivism. *Phenomenology and the Cognitive Sciences*, *11*(4), 467–486.
Kirchhoff, M.D. (2015). Extended cognition & the causal-constitutive fallacy: in search for a diachronic and dynamical conception of constitution. *Philosophy and phenomenological research*.
Krickel, B. (2017). Making sense of interlevel causation in mechanisms from a metaphysical perspective. *Journal for General Philosophy of Science. Zeitschrift fur Allgemeine Wissenschaftstheorie*, *48*(3), 453–468.
Lamb, M., & Chemero, A. (2014). Structure and application of dynamical models in cognitive science. In *Proceedings of the 36th annual meeting of the cognitive science society* (pp. 809–814).
Marr, D. (1982). *Vision: a computational investigation into the human representation of visual information*. San Francisco: W.H. Freeman.
Marsh, K.L., Johnston, L., Richardson, M.J., Schmidt, R.C. (2009). Toward a radically embodied, embedded social psychology. *European Journal of Social Psychology*, *39*(7), 1217–1225.
Maturana, H. (1980a). Biology of cognition. In Maturana, H., & Varela, F. (Eds.) *Autopoiesis and cognition: the realization of the living* (pp. 5–58). Dordrecht: Reidel Publishing Co.
Maturana, H., & Varela, F. (1980b). *Autopoiesis and cognition: the realization of the living*. Boston: D Reidel Publishing.
Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese*, *183*, 283–311.
Raimondi, V. (2014). Social interaction, languaging and the operational conditions for the emergence of observing. *Frontiers in Psychology*, *5*, 899.
Richardson, M.J., & Chemero, A. (2014). Complex dynamical systems and embodiment. In Shapiro, L. (Ed.) *The Routledge handbook of embodied cognition* (pp. 39–50). New York: Routledge.
Rucińska, Z. (2016). Enactive mechanism of make-belief games. In Turner, P., & Tuomas Harviainen, J. (Eds.) *Digital make-believe* (pp. 141–160): Springer International Publishing.
Rudder-Baker, L. (2000). *Persons and bodies*. Cambridge: Cambridge University Press.

Stich, S., & Ravenscroft, I. (1992). What is folk psychology? *Cognition*, *50*, 447–468.
Varela, F., & Thompson, E.E. (1991). *Rosch the embodied mind: cognitive science and human experience*. Cambridge: MIT Press.
Villalobos, M., & Ward, D. (2015). Living systems: autonomy, autopoiesis and enaction. *Philosophy & Technology*, *28*(2), 225–239.
Wilson, A.D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, *4*, 58.
Wimsatt, W. (1997). Aggregativity: reductive heuristics for finding emergence. *Philosophy of Science*, *64*, 372–384.
Zednik, C. (2011). The nature of dynamical explanation*. *Philosophy of Science*, *78*(2), 238–263.
Zednik, C. (2017). Mechanisms in cognitive science. In Glennan, S., & Illari, P. (Eds.) *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 389–400). London: Routledge.