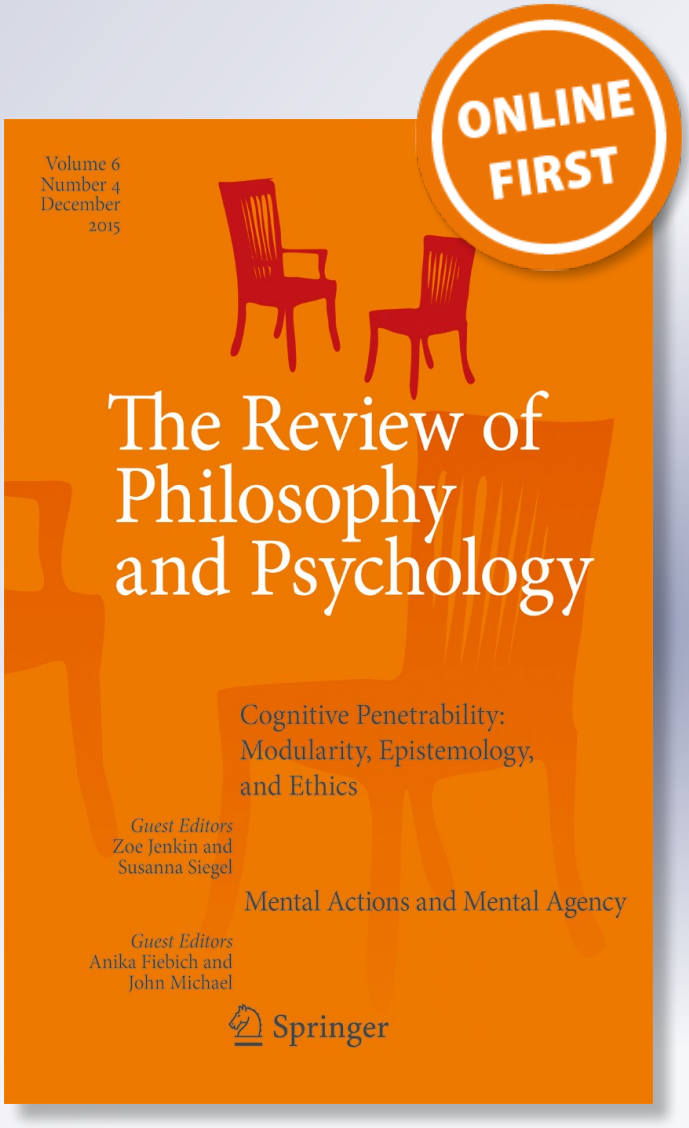
What's so Special About Interaction in Social Cognition?

Julius Schönherr

Review of Philosophy and Psychology

ISSN 1878-5158

Rev.Phil.Psych.
DOI 10.1007/s13164-016-0299-y



Springer

Rev.Phil.Psych.
DOI 10.1007/s13164-016-0299-y

What's so Special About Interaction in Social Cognition?

Julius Schönherr[1]

Abstract Enactivists often defend the following two claims: (*a.*) Successful interactions are not driven and explained by the interactors' ability to mindread (i.e. the ability to attribute beliefs and desires to other agents). And (*b.*) the mechanisms enabling 2^{nd} personal social cognition and those enabling 3^{rd} personal social cognition are distinct. In this paper, I argue that both of these claims are false. With regard to (*a.*) I argue that enactivists fail to provide a plausible alternative to traditional accounts of social cognition in interaction. I examine and reject Hanne De Jaegher's view according to which interaction is "constitutive" for social interaction. Furthermore, I critically discuss Shaun Gallagher's and Daniel Hutto's views according to which social interactions are exclusively driven by low level cognitive mechanisms such as "gaze following" and "emotion detection". Concerning (*b.*), I rely on data from so called "spontaneous response" false belief tasks to show that interactive and observational paradigms require the same "social-cognitive" interpretation.

1 Introduction

For now over 15 years, some researchers in neuroscience, cognitive science, and in philosophy have advocated and defended the idea that our ability to successfully interact with other agents and our ability to understand other agents when we observe them call for different "social cognitive" explanations. However, there is as of now no consensus about how best to describe the "social cognitive" rift that

✉ Julius Schönherr
schoenherrjulius@gmail.com

[1] Department of Philosophy Skinner Building, University of Maryland, College Park, MD 20742 USA

Published online: 26 January 2016

separates interactive and observational contexts. More particularly, on the one hand, researchers disagree about the *kinds* of cognitive mechanisms (e.g. word learning, gaze following, belief-desire attribution, attention allocation) that are recruited differentially in both contexts. On the other hand, there has been widespread disagreement about the *extent* to which cognitive mechanisms are recruited differentially. In this introductory section, I will give an overview over four prominent proposals from the recent literature. Thereafter, I will focus my discussion on one particular interpretation according to which interactive and observational contexts are *categorically* distinct with respect to *belief-desire* attribution.

A number of researchers have construed the interaction–observation divide as a matter of degree. Most prominently, Leonard Schilbach and colleagues' extensive research (e.g. Schilbach 2014; Schilbach et al. 2010, 2013) concerning the relevance of interaction with regard to lower-level social cues such as mutual gaze, joint attention and socially relevant facial expressions has shown that there are distinct neural activation profiles associated with interactive and observational contexts. Firstly, they use interactive eye tracking[1] to show that self-directed facial expressions lead to "a differential increase of neural activity in the ventral portion of medial prefrontal cortex and the (superficial) amygdala, other-directed facial expressions resulted in a differential recruitment of medial and lateral parietal cortex" (Schilbach et al. 2013, 400). Thus, self-directed facial expressions are associated with "emotional and evaluative processing" (Schilbach et al. 2006, 2013). Secondly, when jointly attending to an object the medial prefrontal cortex, and the posterior cingulate cortex are differentially activated (Schilbach et al. 2013, 402). Thirdly, they find distinct patterns of neural activity associated with the different roles agents may have when jointly attending to an object. *Following* someone's gaze directed at an object differentially recruits the medial prefrontal cortex, while *leading* someone's gaze recruits the ventral striatum (Schilbach 2015, Schilbach et al. 2010, 2702).

Interpreting these neural data, they suggest that *leading* gaze may have a rewarding effect and may lead to an increase in motivation (Schilbach et al. 2010, 2013).[2] Along the same lines, they hypothesize that 2^{nd} person interaction is marked by heightened emotional engagement (Schilbach et al. 2013, 396). Furthermore, using a stimulus-response compatibility task,[3] Schilbach et al. (2011) show that gaze shift of an interacting social stimulus influences action control in normal functioning subjects, but not in subjects with high-functioning autism. This indicates that action control in normal functioning subjects is dependent on interactive gaze.

Notably, on Schilbach's account, "social cognitive" differences in interactive and observational situations are wide-ranging as they bear on attention allocation, reward experience, motivation, and action control. It is for these reasons that they speak

[1]This method allows to obtain eye tracking data from "participants inside the MR scanner to make a virtual character's gaze behavior responsive to the participant's gaze in real time" (Schilbach 2014).

[2]Redcay et al. (2013, 435) lends further support to the idea that self-directed gaze in interactions and self-directed gaze from a video replay is associated with distinct neural activity.

[3]In this experiment, neurotypical subjects and subjects with high-functioning autism had to produce spatially congruent and incongruent motor responses in response to either a gaze shift of social stimuli or a shift of an object stimulus.

of a "second person mode" of social cognition that is "fundamentally different" (Schilbach 2014) from third person social cognition. Nevertheless, the differences between these "modes" of social cognition should be understood as a matter of degree.

A second prominent idea concerns the role interaction may play in learning processes. Famously, György Gergely and Gergely Csibra argue for a human-specific learning mechanism which is sensitive to interaction-specific ostensive signals (e.g. eye contact, eyebrow raising) (Gergely 2010; Csibra and Gergely 2006). In these ostensive contexts, according to Csibra and Gergely, the learner is biased to interpret communicative gestures as transmitting generic knowledge about referential kinds (Csibra and Gergely 2009) (rather than just episodic facts). For instance, in one crucial study (Yoon et al. 2008) preverbal infants encode information about an object's *identity* in a communicative context (involving eye contact, and infant directed speech); and they encode information about an object's *location* when such a communicative context is absent.[4]

However, Csibra and Gergely's proposal remains controversial when interpreted as a claim specifically about interaction. On the one hand, although children's sensitivity to different types of information may depend on a context being communicative, it remains to be shown whether it also depends on communication *in interaction*. In Yoon et al.'s experiment (Yoon et al. 2008) a communicative, interactive context was contrasted with a non-communicative, non-interactive context (see footnote 4). Hence, it was not established how infants would have responded had they merely observed a communicative context. Secondly, the role of interaction in learning might be more complex than Csibra and Gergely's model predicts. For instance, Shimpi et al. (2013) find that imitative learning of novel actions is sensitive to toddler-directed ostensive cues *only if the interactor is familiar to the infant*.[5]

A third proposal concerning social cognition in interaction comes from Henrike Moll and Michael Tomasello (Moll et al. 2011) who argue that children often overestimate the amount of knowledge that is shared between her and the person interacted with (Moll et al. 2011, 256). In their study, two-year-olds first played with an adult using two toys. Subsequently, in the "Silent Absence Condition", the adult left the room and stopped the interaction with the child. Then a third toy was introduced to the child in the adult's absence. Then the adult returned. Alternatively, in the "Communicative Absence Condition", the adult left the room but kept communicating with the child in her absence from behind a shelf saying things such as "Oh, how nice! Great! Super!" (see Moll et al. 2011, 256). Moll et al. found that in the Silent Absence Condition all infants knew that the adult had not encountered the third toy which was introduced in the adult's absence. In contrast, in the Communicative Absence Condition children found it significantly harder to tell which object was unknown to the adult. In light of this finding, Moll et al. hypothesize that, in interactive contexts,

[4]In this looking time experiment, infants are shown to be more surprised when an object unexpectedly changes its identity after an actor had pointed to the object in a communicative context. Furthermore, infants are shown to be more surprised when an object unexpectedly changes its location after an agent had grasped the object in a non-communicative context.

[5]In this experiment, 18 month old infants are presented with novel actions (e.g. ringing a doorbell using one's forehead) after a brief warm-up period involving a sorting game. Shimpi found that imitative learning crucially depends on whether the person interacted with later on was familiar from the warm-up period.

young children assume that they share the space around them. However, whether this finding points to a cognitive feature specific to interactive engagement has yet to be empirically determined by introducing more controls in the study. For instance, Moll et al. did not rule out whether children would rely on the 'shared space' assumption when merely observing an interactive situation.

Notably, these three proposals don't explicitly address the role of belief-desire attribution in interaction and observation. A fourth proposal, the one I will focus on in this paper, has been coined "enactivism". Enactivists specifically deny that most interactions involve the attribution of beliefs and desires to other agents (an ability I will call "mindreading"). More specifically, a number of enactivists have argued that mindreading should be relegated to the 3^{rd} personal (i.e. observational) contexts (e.g. Hutto 2004; Gallagher 2001; Reddy 2008). The 2^{nd} personal stance (i.e. the interactive stance) in contrast is, in some important sense, devoid of mindreading. This is not to be understood as a developmental claim alone. Allegedly, interactions between adults don't involve mindreading either. Consequently, enactivists have prided themselves on offering an alternative to the more traditional belief/desire-based approaches to social cognition, i.e. simulation theory (ST)[6] and theory theory (TT)[7] (In what follows, I will refer to theorists who defend one of those theories collectively as "ToMers"). If the enactivist's assessment should turn out to be correct, then ToMers are thoroughly undermined, because, of course, they aspire to accurately capture the cognitive mechanisms underlying real life interactions (see Carruthers 2009, 167).

In short, among other things, enactivists have defended the following two ideas:

Core Thesis. Successful interactions are neither driven nor explained by the interactors' ability to mindread.
Distinctness. The mechanisms enabling 2^{nd} personal social cognition and those enabling 3^{rd} personal social cognition are distinct.

In Section 2, I will clarify what is meant by 'interaction'. Next (Sections 3 and 4), I will examine two enactivist defenses of **Core Thesis**. The first defense was developed by Hanne De Jaegher and Ezequiel Di Paolo who argue that interactions cannot be explained in terms of mindreading, because, quite generally, interactions cannot be explained in terms of individual agents' contributions to interactions. Next, in Section 4, I will critically discuss Shaun Gallagher's and David Hutto's account of **Core Thesis**. Both argue that lower level cognitive mechanisms not involving mindreading suffice to navigate interactions. In Section 5, I will argue against Gallagher's and Hutto's view concerning **Distinctness.**

Before starting my discussion proper I'd like to motivate my skeptical stance towards enactivism by discussing two general worries. These worries will cast initial doubt on the idea that social cognition in interactive and observational contexts is fundamentally distinct.

[6] According to simulation theory, attribution of mental states to other agents is achieved using one's own mental states to simulate the mental states of other agents.

[7] According to theory theory, the attribution of mental states to other agents is achieved through the application of a 'theory'.

First, in the real world, the boundaries between 2^{nd} and 3^{rd} personal contexts simply aren't clear-cut enough in order to merit such a sharp theoretical distinction. Suppose, for instance, a child, Johnny, is interacting with his brother in order to plot something against their sister, Mary, who plays with her toys in the other end of the room. Suppose Johnny now turns to Mary and starts interacting with her. The enactivist might say that Johnny first had a purely observational attitude towards his sister; thereafter he adopted an interactive attitude towards her. But how plausible is this? In the real world, interactions with and observations of others are tightly interwoven; so tightly indeed that it would seem surprising if distinct *theories* were to apply to both contexts. Johnny might observe his sister for a few seconds, then interact with her for a minute, then turn back to his brother observing her again. In the face of such a tightly knit juxtaposition of interactive and observational contexts we might generally tame our expectations concerning hard and fast distinctions between both contexts.

Second, many inferences that can be drawn specifically *in* interactions can likewise be drawn when *observing others interact*. Let me give an instructive example from Stephen Butterfill (see Butterfill 2013): He argues that in interactive situations involving joint actions, it is especially easy to correctly attribute intentions to the other agent; this is simply because joint action often requires sharing of intentions. Suppose I intend to put a stroller on a bus and you help me carry out my plan. If our joint action is to be successful you should, by default, also intend to put the stroller on the bus. If you, say, intend to flip it over, or take its wheels off we won't succeed in carrying out my plan. Butterfill writes that in these situations interactors "may be in a position to know that the goals of her target's actions will be the goals of her own actions" (Butterfill 2013, 22). Hence, interactions of this form make effortless attributions of goals to others possible. However, at the same time Butterfill recognizes that observers could also acquire such knowledge. While interactors can rely on the "my-goal-is-her-goal" inference, observers could respectively rely on the "her-goal-is-his-goal" inference (Butterfill 2013, 20). Just as I am in a position to know your intentions when we are carrying out a joint action, so can an onlooker know your intentions when she sees us carrying out a joint action (given she knows my intentions). Interactors do not enjoy a principled privilege. Therefore, at least in this case, alleged genuinely "interactive" features of social interactions can, in principal, be exploited from a third person perspective. Of course, if we put a stroller on a bus I will usually be more motivated to know what your intentions are. Furthermore, I will pay closer attention to what you are doing than as if I were just watching the scene unfold. Therefore, the fact that there is no principled advantage interactors enjoy leaves untouched Schilbach's claims about the role of motivation, attention, and reward feelings in interactions (see above).[8]

[8]Furthermore, research concerning social cognition in high-functioning autism indicates that there is *some* difference between social cognition in interactive and observational contexts. Schilbach et al. (2013) hypothesize that it is specifically social cognition in interaction which may be impaired in high-functioning autism.

2 Interaction—Paradigms Without a Definition

ToMers and enactivists disagree about whether normal subjects (need to)[9] rely on mindreading in order to successfully navigate social interaction. But what are interactions? In her article "Embodied cognition and mindreading" (Spaulding 2010) Shannon Spaulding gives us a concise example of a prototypical interaction:

Suppose Jack and Jill are sitting in a coffee shop; both are doing some work on their respective computers when suddenly Jack starts asking Jill questions such as "What are you working on?", "Where are you from?" etc.. When Jill gives only cursory answers such as "philosophy", Jack responds "Oh, I bet you are really deep" to which Jill just responds "sure". This goes on for a little while but, ultimately, when he realizes that Jill won't reciprocate the way he'd like, Jack lets Jill off the hook and they both go about their work.

If ToMers and enactivists disagree about what enables agents to successfully interact, they surely disagree about the kind of situation just described. ToMers will most likely analyze this situation along the following lines: Jill believes that Jack believes that she is interested in him. She also believes that Jack's belief is false, because, in fact, she is not interested in him. Jack initially believes (or hopes) that Jill has a desire to talk to him, but when Jill keeps giving curt answers he finally realizes that this belief was false (see Spaulding 2010). Hence, ToMers believe that what drives social interaction is mindreading. Enactivists reject this interpretation.

Definitions of the relevant terms '2^{nd} personal stance', '3^{rd} personal stance', 'interaction', and 'observation' etc. are hard to find in the literature.[10] However, I believe that, ultimately, it is not necessary to provide such definitions, which would

[9]Recent studies concerning the automaticity of mindreading (e.g. Qureshi et al. 2010; Schneider et al. 2014) seem to indicate that others' mental states may be computed and attributed to others even in situations in which this is not at all necessary. Therefore, the claim that mindreading is necessary for social interaction needs to be distinguished from the claim that mindreading is, in fact, employed.

[10]Notably, De Jaegher and Di Paolo (2007) is an exception. They give the following definition of "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 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, 493).

Following an interpretation by Herschbach (2012), "coupling" amounts to the coordinated mutual dependence of the behavior of several subjects. Furthermore, coupling can be said to be "*regulated*" if "engaging in motivated changes to the constraints or parameters that influence the coupling" (Herschbach 2012). One worry concerning De Jaegher's definition is that it might not accurately distinguish interaction from mere coordination. Arguably, coordinated action also requires regulated coupling between autonomous agents.

be called for if there were vast disagreement about which situations are interactive, observational etc.. But this is not the case. Philosophers and cognitive scientists by and large agree about which cases they are disagreeing about. The disagreement is about the *correct analysis* of the relevant cases, not about their identity.[11]

It is worth noting that the interaction/observation distinction cannot be sufficiently identified using a grammatical criterion. The different stances don't map onto the grammatical distinction between the use of the personal pronouns "you" (for the interactive stance) and "he", "she", "it", and "they" (for the observational stance). Simply put, there are interactions in which we don't use any of these pronouns. Second, the use of these pronouns doesn't tell us anything about the nature of the respective stances. They could, if anything, just provide a marker.

3 The Constitutivist Account of "Core Thesis"

In a series of articles Hanne De Jaegher has developed the view that interaction is constitutive of social cognition[12] (e.g. De Jaegher et al. 2010; De Jaegher and Di Paolo 2007); i.e. interaction is an essential proper part of social cognition. This account is intended to stand in opposition to traditional ToMistic views according to which social cognition is reducible "to the workings of individual cognitive mechanisms" (De Jaegher et al. 2010). De Jaegher argues that "interactive processes [...] complement and even replace individual mechanisms" (De Jaegher et al. 2010). Her view can be summarized as follows:

> **Constitutivism about Interaction.** Central features of social cognition in interactions cannot be explained solely in terms of each interactor's contributions to these interactions (e.g. their behavior and their mental states). Rather, interaction is explanatorily basic for social cognition.

Note that De Jaegher's claim is quite radical in that she does not merely hold that features of interaction can *causally* influence an individual's cognitive processing (which would be uncontroversial).

A general problem with her approach, pointed out by Herschbach (2012), is that her constitution claim involves a category mistake. Herschbach asks "[i]n what sense could a social interaction be a constitutive element of a *neural mechanism*?". And

[11] Note that other related discussions may well benefit from definitions of these terms. Categorization of the relevant cases is far less obvious when, say, comparing interaction to cooperation. After all, it is not intuitively clear which cases exemplify cooperation and which ones exemplify interaction. It is just that in the present discussion these definitions are not necessary.

[12] "[S]ocial cognition", in this context, is defined as a "[g]eneral term used to describe different forms of cognition, about, or actions in regard to, agents or groups of agents, their intnetions, emotions actions and so on, particulary in terms of their relation other agents and the self" (De Jaegher et al. 2010).

he continues "[i]f a constitutive element is understood as a 'part of the phenomenon' itself, this statement would involve a substantial confusion between levels of organization" (Herschbach 2012, 477). The worry, put more colloquially, is that interaction is something that goes on *between subjects* while neural processes are something that happen *inside one subject*. One may worry that this critical assessment overextends in that it would amount to a *general* criticism of the extended mind thesis (i.e. the claim that cognition is not confined to what's going on beneath the skull). Hence, rejecting De Jaegher's claim would require a more thorough treatment of the extended mind literature.

A more immediate problem for De Jaegher's constitution claim is that her examples don't unambiguously support her case. For instance, she relies on a "perceptual crossing" experiment (De Jaegher et al. 2010) conducted by Auvray et al. (2009). In this experiment, two blindfolded participants interact with each other by moving an avatar along a one dimensional strip. For each player's avatar there is also a shadow that replicates the avatar's movements. Furthermore, along the strip there is an additional static object. Hence, each player can encounter three different objects: The other player's avatar, the other player's shadow, and the static object. When a player encounters any object she receives sensory feedback.

For all three objects, the sensory feedback is identical; hence, a player cannot distinguish between the different objects encountered by relying on sensory information alone. Crucially, when two avatars meet, *both* players receive sensory feedback. When an avatar encounters a shadow only the player whose avatar it is receives feedback.

De Jaegher notes that "in such an impoverished sensory situation, participants find each other and concentrate their mouse clicks on each other's sensors (65.9 % of clicks) and not on the identically moving, but non-contingent shadow objects (23 %)" (De Jaegher et al. 2010, 444). The players' "finding each other" is explained by their behavior when they encounter an object. Upon receiving sensory feedback, a player tends to reverse her direction. When an avatar encounters a shadow, the avatar reverses direction, but the shadow does not. When two avatars meet, they both tend to reverse direction and start oscillating around each other. According to De Jaegher, this experiment provides evidence that the agents' finding each other can't be explained by each player's contribution. This is because each player is inept to even distinguish a shadow from an avatar.

De Jaegher's interpretation can be resisted. The perceptual crossing experiment does not establish that interaction is basic or *constitutive* of social cognition. Granted the experiment *does* show that not all features of the interaction can be explained by just looking at *one* player's contribution. However, everything that happens in Auvray's experiment is entirely predictable if we take into account *both* players' contributions. The interaction effect De Jaegher describes is fully determined by the pattern of sensory feedback that each player receives in conjunction with their individual strategies (i.e. reversing the avatar's direction upon receiving sensory feedback). The idea, however, that all facts about interactions cannot always be explained solely in terms of *one* interactor's contribution is not very controversial. Suppose, for instance, you and I want to put a stroller on a bus. What explains our success

in completing this task? Surely, a satisfactory explanation would need to appeal to both of our goals; it would need to take into account that your actions (say, the speed with which you lift the stroller) have an effect on what I do (say, lifting the stroller with equal speed). ToMers can embrace the idea that each interactor's actions have a *causal* effect on the respective other's cognition. However, what these examples do not show is that interaction is *constitutive* for social cognition. I conclude that the relevant data in support of De Jaegher's view *can* be explained within an individualistic paradigm.

In this section, I argued against De Jaegher's view that interaction cannot be understood in terms of individual agents' cognitive mechanisms. My main line of reasoning was that the constitutivist approach lacks convincing examples. In the next section, I will focus on the more moderate enactivist theory developed by Shaun Gallagher and Daniel Hutto who argue that, although interaction is not "basic" in the aforementioned sense, successful interaction does not require mindreading.

4 Shaun Gallagher's and Daniel Hutto's Account of "Core Thesis"

Shaun Gallagher and Daniel Hutto defend a type of two-systems account of social cognition.

System 1 operates fast and unconscious. It does not involve mindreading, but, rather, exhaustively recruits lower-level mechanisms which Gallagher and Hutto label "primary" and "secondary intersubjectivity" (PIS and SIS). PIS enables agents to interact with *one another*; SIS enables agents to adopt a shared perspective with regard to *the world*. PIS comprises cognitive mechanisms such as "gaze following", "emotion detection", and understanding of goal-directed actions. SIS comprises cognitive mechanisms such as "joint attention" and the ability to understand others' emotionally valenced attitudes towards an object or a situation (see Gallagher 2001, 2008, 2012; Hutto 2004).

System 2—the mindreading system—is slow, non-modular and solely consciously employed. Think, for instance, of the reasoning underlying Sherlock Holmes's painstaking reconstruction of the murderer's motive. Crucially, such conscious reasoning about mental states is not fast enough to guide and direct interaction. Typically, Gallagher and Hutto maintain, in interactions there is no time for a slow and cognitively costly reconstruction of the other agent's mental states. Therefore, agents have to rely on low level cognitive mechanisms PIS & SIS.

Furthermore, mindreading is supposed to be constitutively 3^{rd} personal (Hutto 2004). Hutto explains that "[w]e ascribe causally efficacious inner mental states to them [other agents] for the purpose of prediction, explanation, and control" (Hutto 2004, 549). This amounts to viewing them as "foreign bodies" (ibid., 549) and, thereby, taking a spectatorial stance towards them. In contrast, when we interact with other agents we rely on more basic forms of "primary" and "secondary intersubjectivity" (see Gallagher 2012, 2001, 89; Hutto 2004, 550).

Paradigmatically, we take a 3^{rd} personal stance towards other people when their actions seem unfamiliar and atypical to us (see Gallagher 2001, 92; Hutto 2004).

In such cases, we theorize about others' beliefs and desires in trying to explain their actions. However, when everything goes as usual mentalizing is unnecessary.[13]

System 2 reasoning about mental states is essentially conscious. Therefore, few will deny its existence, and I won't spend time on this part of the enactivist theory. Disagreement arises with regard to the enactivist's system 1. Is it true that the unconscious cognitive processes which guide and regulate swift social interaction do not employ mindreading? Do system 1 mechanisms SIS and PIS provide sufficient cognitive resources to explain successful social interaction?

Let me discuss what I take to be the two most pertinent and contentious issues. First, infants reliably pass interactive, non-verbal false-belief tasks at 18 months of age and younger. It is clear that at this young age, infants couldn't possibly *consciously* reason about false beliefs, or rely on narratives to guide their understanding. Hence, the enactivist is called upon to give an ersatz-explanation that does not rely on mental state attribution. Second, Hutto and Gallagher adduce a principled argument for why mentalizing *cannot* drive social interactions. This argument states that, by ToMers own lights, mentalizing is used to "predict and explain" others' behavior. However, predicting and explaining behavior couldn't possibly be an unconscious process. If mindreading is a conscious process, then it could not underlie interaction (for reasons stated above). I will discuss both issues in turn.

Firstly, in a study by Buttelmann et al. (2009), 18 months old infants succeed in helping an adult retrieve an object from a box, while, according to the standard interpretation, taking into account the adult's false belief about the object's location. In the experiment, an infant watches how an adult sees an object being placed into one of two boxes (box A). Then, in the false-belief condition, the object is moved from box A to a different box (box B) in the adult's absence. When the adult finally tries to retrieve the object from box A (due to her false belief) the child helps the adult, leading her to box B which contains the object. The infant, however, only helps the adult retrieve the object in the false-belief condition. In the true-belief condition in which the adult knows the true location of the object and yet still opens the empty box, the child helps the adult open the empty box assuming that she must have some other reason to open it.

According to the mentalizing interpretation of the active-helping study (which Gallagher rejects), the infant understands that, in the false-belief condition, the adult *believes* that the object is in box A, and that she *wants* this object. This is what motivates the infant to help. According to a different, non-mentalizing interpretation (usually labeled "the behavior rule interpretation"), the infant knows a rule such as "people look for objects where they last saw them". Rules such as this one are meant to enable the infant to, say, distinguish situations in which the adult looks for an object from situations in which she does not look for it. This, in turn, is important for knowing when to help and when not to help.

[13] If, in a given situation, these low level cognitive tools don't suffice, according to Hutto, narratives help us become familiar with social situations (e.g. Hutto 2009).

Gallagher rejects both of these interpretations, arguing that there is a distinctively enactive way of viewing these findings. He states:

> [...] the fact that the infant knows either that the agent has been in a position to see the switch or not, plus the agent's behavior with respect to A [...], is enough to specify the difference in the agent's intention. For the infant, that signals a difference in affordance, i.e., a difference in how the infant can act, and thereby interact with the agent. The infant does not have to make inferences to mental states since all of the information needed to understand the other and to interact is already available in what the infant has seen of the situation (Gallagher 2012, 201).

And

> The phenomenological-enactive approach provides an alternative to both the ToM and behavioral interpretations (Gallagher 2012, 202).

Notably, in this passage, Gallagher focuses on knowing "intention[s]" (and not on knowing behavior). Accordingly, the enactivist may depart from a behavior rules account of social cognition by relying on intention rules: "people *intend* to look for objects where they last saw them". This interpretation, very much in the behavior-rule spirit, introduces a further complication: Once the infant knows the adult's intentions, she then has to employ the additional rule "people tend to do what they intend". Therefore, by putting intentions in the focus of analysis, the enactivist cannot hope to get around a behavior rule which maps intentions to actions. Furthermore, it is not clear what the motivation for such an 'intention-rule' could be. One of the attractions of behavior rules is their alleged parsimony (they don't involve mentalizing of any sort). Intention-rules, on the other hand, *do* involve mentalizing (they involve *intention* attribution); hence such rules would be *less* parsimonious, and, therefore, we'd be owed an account concerning the benefits of such rules.

A second, distinctively enactive perspective concerns the explanatory role of agents' possibilities for action. Standardly, ToMers hold that the infant's action possibilities (e.g. the possibility to help the adult open the box) are *grounded in* an understanding of the situation; in an understanding that the adult wants the object and that she has a false belief about its location. Alternatively, avowed enactivists sometimes hold that the direction of explanation should be reversed: Action possibilities sometimes ground how objects and situations are represented. This line of reasoning is famously adopted by Alva Noë who argues that certain properties of perceptual content are constituted by sensory-motor relations. His most thoroughly discussed example concerns the perception of a tomato (see Noë 2008). He starts with the following observation: Looking at a tomato we can only literally see one side of it (the side facing us). Nevertheless, we *perceive* tomatoes as three dimensional objects. We perceive it as an object which has some hidden sides. According to Noë, the perceptual content of a tomato as a three dimensional object is *constituted by* or *grounded in* the "availab[ility] to perception through appropriate movement" (Noë 2008, 16). The tomato's hidden sides are present in perception, *because* upon perceiving one side one has motor access to a visual representation of its hidden sides.

This line of reasoning is reflected in the following quote from Gallagher: "[I]nfants understand others *in terms of how they can interact with them*" (Italics by the author) (Gallagher 2012). Hence, understanding others' mental states may be grounded in possible ways to act and interact with the agent. However, without taking a stance on Noë's view on perception, this line of reasoning is hardly plausible for the relevant social cognition cases. To see this, reconsider the active helping study. Suppose that the infant's understanding of the adult's beliefs and desires were based on her grasping that it is appropriate to help in one situation but not in the other. What, then, explains the child's sensitivity to situations in which helping is (or is not) appropriate? Surely, it cannot be the attribution of beliefs and desires. But grasping action possibilities cannot be *bare* either, simply because different situations afford different actions; and first agents need to understand a situation in order to know which actions are afforded.

Noë's perception-based examples and the "social-cognitive" paradigms are disanalogous in an important sense. On the one hand, we have lots of experience perceiving, handling, and modifying objects. This is what supposedly grounds sensory-motor expectations. We know that objects such as tomatoes will reveal hidden sides when we go around them and when we move them in our hands, because we have seen this happen many times before. It is not clear what the relevant prior experience in the active helping study would be. Surely, we have *vastly* more experience discovering hidden sides of three dimensional objects than we have with others' false beliefs; especially at 18 months of age. The situation in which an object is moved from one box to another while the adult is absent is comparatively unique for the child.

Yet another line of argument frequently adopted by enactivists relates social cognition to "direct perception" (Gallagher 2008, 536). According to Gallagher, we "have a direct perceptual grasp of the other person's intentions, feelings, etc." (Gallagher 2008). Intention attributions are therefore not *mediated* by either a theory or a behavior rule, which means that "there is no problem of other minds" (Gallagher 2008). Because mental states can be perceived directly, there is simply no need for any intermediate cognitive mechanisms.

The 'immediate perception' view is problematic if intended to provide an *alternative* to mentalizing accounts of social cognition. This is because questions concerning the contents of perception and questions concerning the underlying mechanisms of social cognition should be kept distinct. The basic argument is this: If perceptual content is conceptual at least in some cases, then it is a live option for theory theorists to argue that the conceptual outputs of a mindreading module can be a constitutive part of perceptual states (see Carruthers 2015). Of course, this would not provide an immediate answer to the question whether *mental state concepts* can figure in perception. However, it would guarantee the in-principle compatibility of theory theory and a direct perception account of mental states.

Let me give two arguments in favor of the view that concepts can be part of perceptual states. The first argument is phenomenological in nature, the second draws on the tight interplay between perception and cognition.

In a recent article, Carruthers (see Carruthers 2015) suggests that when we see something, S, as an instance of a kind, K, the concept that represents K is "bound into" the perception of S (Carruthers 2015, 6). For an illustration, think of perceiving a cloud. You stare at a cloud when all of the sudden you realize that it looks face-like (or, say, wardrobe-like); i.e. you see the cloud *as* a face, or *as* a wardrobe. In such cases, Carruthers argues, the concept FACE is bound into the perception of the cloud. Along the same lines, theory theorists could argue that the conceptual outputs of a mindreading theory module can be "bound into the contents of the perceptual states that provide the basis for its interpretations" (Carruthers 2015, 7). Such examples are persuasive, because it is hardly plausible that the mental representations FACE, or WARDROBE are essentially non-conceptual.

Secondly, a growing body of literature indicates that there is a tight interplay between conceptual knowledge and visual perception. One plausible explanation of this interplay is that perception just is conceptual. Evidence comes mainly from research concerning links between color perception and color concepts (e.g. Thierry et al. 2009; Winawer et al. 2007; Daoutis et al. 2006). One experiment by Daoutis et al.'s (2006) involved 4–7 year old children from either England or Kwanyama (Namibia). The crucial difference between both groups was that the Kwanyama don't have distinct color terms for the colors blue and green, blue and purple, and red and pink. In the experiment, the children had to find a target color in an array of color patches which contained patches of either the target color or distractor colors. The distractor colors were designed to be either cross-category for English speakers, and within-category for Kwanyama speakers; or, in a second condition, cross category for both groups. Daoutis et al. found that within-category search was faster for the English speakers. This effect did not hold for the Kwanyama speakers (for whom there was no within-category condition). One attractive (but not the only available) interpretation of the data is that color concepts form a part of color perception. This would explain why differences in conceptual knowledge predict performance in visual search tasks.

Carruthers considers an alternative explanation of these findings according to which "concept acquisition permanently "warps" the processing that takes place in midlevel visual areas" (Carruthers 2015, 9). However, he argues that long-term "warping" is unlikely, because interaction effects between color perception and color concepts are highly sensitive to online interference effects. Typically the concept-based performance differences in these tasks go away under cognitive load. These arguments don't conclusively settle whether or not perception itself is conceptual (or whether conceptual knowledge merely has causal effects on perception). But my goal is more moderate. I showed that ToMistic accounts of social cognition are in principle able to embrace a direct perception account of social cognition. It is at the very least a live option for ToMers to hold that perceived mental states could be the result of a complex interplay between conceptual mindreading systems on the one hand and perceptual systems on the other.

Let me now go on to discuss enactivist claims concerning the role of prediction and explanation of behavior. Enactivists have argued that folk-psychological attributions

of beliefs and desires serve to *causally predict and explain* behavior (Gallagher 2001, 102; Gallagher 2012; Hutto 2004, 549). According to enactivists, ToMers share this view.

> The claim that what one is doing when mindreading is *explaining* or *predicting* the other person's action in terms of mental states, however, is not *my* claim. It's a claim that is pervasive in the ToM literature (Gallagher 2012, 205).

Similarly, Hutto writes

> it is also generally assumed [by ToMers] that we are normally at theoretical remove from others such that we are always ascribing causally efficacious mental states to them for the purpose of prediction, explanation and control (Hutto 2004, 548).

Allegedly, this particular mode of understanding others leads to "estrangement", and, therefore, it cannot serve as the right model for understanding others in interactive contexts. Hutto argues that predicting and explaining others' actions is only necessary when actions are unfamiliar to us. In most circumstances, however, "we already know what to expect from others and they know what to expect from us in familiar social circumstances" (ibid., 558).[14]

Now, "know[ing] what to expect" does not absolve us from *predicting* what others do. Suppose, for instance, you bump into somebody in the hallway whom you want to pass. Suppose a convention exists according to which, in these situations, both people step to their respective right. Surely, in this case, you know what to expect from the other person: she will take a step to her right. You take a step to your right and, hence, you both succeed in passing each other. You both knew what to expect, because you both knew the pertinent rule for such situations. However, the fact that this coordination problem was particularly effortlessly and easily solvable does not mean that you didn't have to predict what the other person would do. You predicted that she would take a step to the right and that is why you stepped to the right. Of course, this does not entail that the enactivist's analysis is wrong. In fact, there is little reason to assume that solving the hallway problem involves mindreading. All it shows is that enactivists will also have to appeal to the prediction of others' behavior at some level in their theory. In the hallway case, the prediction might have been facilitated by the existence of a social convention to step to the right, which may have obviated the need to mindread. Social predictions and explanations can be

[14] According to Hutto, one reason to favor narrative-based accounts over ToMistic accounts is its phenomenological accuracy. We simply don't go around consciously calculating others' beliefs and desires all the time. However, it is doubtful that narrative-based models fare better with regard their phenomenological accuracy. As it is, we also don't go around recalling stories that might fit a particular interactive situation. Understanding others is often *entirely* effortless. Therefore, any theory about social cognition which gives lots of weight to phenomenological 1^{st} person data will have to refrain from positing *any* explanatory mechanism. This, however, seems implausible. As John Michael argues, surely, in understanding others, interpretation has to happen somewhere (see Michael 2011, 562).

accomplished in various ways (e.g. through mental state attributions, social conventions, or behavior guiding rules). They don't always involve mindreading. However, the view that ordinary social circumstances don't require any predictions is flawed and ToMers' positions cannot be ruled out on those grounds alone.

A similar argument can be given for 'explanation'. Actions are often ambiguous; one and the same physical action can mean different things and can be interpreted in various ways. Reconsider, for instance, the active helping study. When the agent comes back and tries to open a box, there are several things she could be interpreted as doing. She may be trying to open the box, lift the box, or break off the handle. Surely, an adequate understanding requires ruling out some of these possibilities. Ascribing causally efficacious beliefs and desires is one way to reach adequate understanding. According to this model, the child understands what the agent does when she knows that the adult wants the object and that she has a false belief about its location.

Enactivists point out that disambiguation does not always require mentalizing. Rather, understanding is achieved by certain behavioral scripts and lower level cognitive mechanisms. For instance, suppose you stand at the register in the super market. The person behind the register reaches towards you. The display reads $10.53. The appropriate action in this context is to hand her the money. How did you know that this would be the appropriate thing to do? One possibility is that you understood that she *wanted* money from you (and you owe the money). An alternative explanation is that acting in this way was just demanded by the situational setting. Whichever description turns out to be right, it is clear that there needs to be some disambiguating explanation of why the person behind the register acted the way she did. Hence, the ToMistic view cannot be ruled out on the grounds that they provide some such explanation.

In this section, I argued that the enactivist defense of **Core Thesis** does not provide a genuine alternative to more conventional accounts of social cognition. On my interpretation, Gallagher and Hutto's views are close to a behavior rule account of social cognition which is then combined with a direction-perception account of mental states. Furthermore, I argued that mentalizing accounts of social cognition cannot be ruled out merely on the grounds that they involve explanations and predictions of behavior.

5 Gallagher's and Hutto's Account of "Distinctness"

Suppose the enactivist were right in that paradigmatic interactions are free of mindreading. In this case, as I will now go on to show, she is forced to give an enactive analysis of some entirely 3^{rd} personal false belief paradigms (hence, undermining the theoretical distinctness between 2^{nd} and 3^{rd} personal paradigms).

Evidence comes from so-called "spontaneous response" tasks (e.g. Onishi and Baillargeon 2005; Surian et al. 2007; Woodward et al. 2009). In these tasks, children's understanding of others' false beliefs is inferred from "behaviors they spontaneously produce as they observe a scene unfold" (Baillargeon et al. 2010). There are two types of spontaneous tasks. On the one hand there are violation-of-expectation praradigms

which exploit the fact that an infant will look longer at an agent or a scene, if her actions don't match the infant's expectations. On the other hand, there are anticipatory looking tasks which exploit the fact that infants will look in the direction of a location in which they anticipate others to act. Anticipatory looking can be sensitive to (false) belief attribution, because the infant predicts the agents' actions based on belief attribution. Importantly, spontaneous response tasks are observational paradigms in which the infant merely watches a certain scene unfold.

For instance, in a violation-of-expectation paradigm, Onishi and Baillargeon (2005) found that 15-month-olds have an understanding of false beliefs. In this experiment, infants were first familiarized with a toy that stands between a green and a yellow box and which is then hidden in the green box. Next, the agent reached inside the green box to retrieve the toy. Next followed a belief induction faze. In the false belief condition the toy was moved from the green to the yellow box while the adult was absent. When the adult reached for the box where she didn't believe the toy to be, infants looked reliably longer than when the adult reached for the ball in a location incongruent with her false belief about the ball's location; hence, taking into account the adult's false belief, the infant expected her to look for the toy where she falsely believed it to be.

It is clear that enactivists cannot readily embrace the false-belief-tracking interpretation. They believe that belief-desire attributions are the product of *conscious* reasoning. Violation-of-expectation paradigms conducted with 15-month-old infants are not the purview of enactive system 2 mindreading. Furthermore, it would seem quite ad hoc to suppose that Baillargeon's violation-of-expectation paradigm on the one hand and Buttelmann's active helping paradigm are in some basic theoretical way distinct. The only motivation for this view would be the defense of **Distinctness.** But, as I said, this seems ad hoc. Hence, the enactivist needs to explain the violation-of-expectation findings relying on enactivist tools (e.g. in terms of primary and secondary intersubjectivity). Though there is nothing interactive about this paradigm; it is entirely observational.

Therefore, if the enactivist is right in that belief/desire attributions are the result of effortful conscious thought processes, she has to admit that, at least in some cases, enactive social understanding is 3^{rd} personal. This is because infants couldn't possibly consciously reasons about others' beliefs and desires. If, however, the enactivist admits that Baillargeon's paradigm does involve mind-reading, then she also has to be comfortable with the idea that mindreading is a largely effortless, unconscious process.

6 Conclusion

Although it is plausible that social cognition evolved in order to navigate social interactions (Carruthers 2009, 167), hard and fast cognitive differences between interactive (i.e. 2^{nd} personal), and observational (i.e. 3^{rd} personal) situations prove not be supported by the evidence. I have discussed four claims in support of this claim: First, in real world scenarios interactive and observational paradigms are tightly interwoven. Second, certain allegedly interaction-specific inferences have 3^{rd}

personal counterparts and can therefore also be drawn from an observational perspective. Third, De Jaegher's claim that interaction *constitutes* social cognition is untenable. Fourth, the enactivist idea to relegate mindreading to 3^{rd} personal contexts is implausible.

All told, distinguishing 2^{nd} and 3^{rd} personal contexts based whether they involve mindreading, understood as the attribution of beliefs and desires to other agents, is not plausible. However, social cognition in both contexts may still be distinct in less extreme ways. The growing body of research on the automaticity and spontaneity of mindreading (see Qureshi et al. 2010; Surtees and Apperly 2012, Schneider et al. 2014) may shed further light on subtle issues concerning the exact conditions under which mindreading is employed. Moreover, thorough research by Schilbach and colleagues show that interactions involve distinct patterns of neural activation which is associated with motivational, attentional, and reward related "social cognitive" differences. The evalution of this research would be a further step towards fully understanding whether there is something special about interaction in social cognition.

References

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

Baillargeon, R., R. Scott, and Z. He. 2010. False-belief understanding in infants. *Trends in Cognitive Sciences* 14(3): 110–118.

Buttelmann, D., M. Carpenter, and M. Tomasello. 2009. Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112(2): 337–342.

Butterfill, S.A. 2013. Interacting mindreaders. *Philosophical Studies* 165(3): 841–863.

Carruthers, P. 2009. How we knowour own minds: the relationship between mindreading and metacognition. *Behavioral and brain sciences* 32(02): 121–138.

Carruthers, P. 2015. Perceiving mental states. *Consciousness and cognition* 36: 498–507.

Csibra, G., and G. Gergely. 2006. Social learning and socialcognition: the case for pedagogy. *Processes of change in brain and cognitive development. Attention and performance* XXI: 249–274.

Csibra, G., and G. Gergely. 2009. Natural pedagogy. *Trends in cognitive sciences* 13(4): 148–153.

Daoutis, C. A., A. Franklin, A. Riddett, and A. Davies Clifford. 2006. Categorical effects in children's colour search: a cross-linguistic comparison. *British Journal of Developmental Psychology* 24(2): 373–400.

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

De Jaegher, H., E. Di Paolo, and S. Gallagher. 2010. Can social interaction constitute socialcognition?. *Trends in cognitive sciences* 14(10): 441–447.

Gallagher, S. 2001. *The practice of mind. theory, simulation or primary interaction?* Vol. 8, No. 5–7. Imprint Academic.

Gallagher, S. 2008. Inference or interaction: socialcognition without precursors. *Philosophical Explorations* 11(3): 163–174.

Gallagher, S. 2012. Indefense of phenomenological approaches to social cognition: interacting with the critics. *Review of Philosophy and Psychology* 3(2): 187–212.

Gergely, G. 2010. Kinds of agents. *The Wiley-Blackwell Handbook of Childhood Cognitive Development* 22: 76.

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.

Hutto, D. 2009. Folk psychology as narrative practice. *Journal of Consciousness Studies* 16(6–8): 9–39.

Hutto, D.D. 2004. The limits of spectatorial folk psychology. *Mind and Language* 19(5): 548–573.

Michael, J. 2011. Interactionism and mindreading. *Review of Philosophy and Psychology* 2(3): 559–578.
Moll, H., M. Carpenter, and M. Tomasello. 2011. Social engagement leads2-year-olds to overestimate others knowledge. *Infancy* 16(3): 248–265.
Noë, A. 2008. Precis of action in perception. *Philosophy and phenomenological research* 76(3): 660–665.
Onishi, K.H., and R. Baillargeon. 2005. Do 15-month-old infants understand falsebeliefs? *Science* 308(5719): 255–258.
Qureshi, A.W., I.A. Apperly, D. Samson, and Cognition. 2010. Executive function is necessary for perspective selection, notlevel-1 visual perspective calculation: evidence from a dual-task study of adults 117(2): 230–236.
Redcay, E., K. Rice, and R. Saxe. 2013. Interaction versus observation: a finer look at this distinction and its importance to autism. *Behavioral and Brain Sciences* 36(04): 435–435.
Reddy, V. 2008. *How infants know minds*. Cambridge: Harvard University Press.
Schilbach, L. 2014. On the relationship of online andoffline social cognition. *Frontiers in human neuroscience* 8.
Schilbach, L. 2015. Eye to eye, face to face and brain to brain: novel approaches to studythe behavioral dynamics and neural mechanisms of social interactions. *Current Opinion in Behavioral Sciences* 3: 130–135.
Schilbach, L., S.B. Eickhoff, E.C. Cieslik, B. Kuzmanovic, and K. Vogeley. 2011. Shall we do this together? social gaze influences action control in acomparison group, but not in individuals with high-functioning autism. *Autism*: 1362361311409258.
Schilbach, L., B. Timmermans, V. Reddy, A. Costall, G. Bente, and T. Schlicht. 2013. Toward a second-person neuroscience. *Behavioral and Brain Sciences* 36(04): 393–414.
Schilbach, L., M. Wilms, S.B. Eickhoff, S. Romanzetti, R. Tepest, and G. Bente. 2010. Mindsmade for sharing: initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience* 22(12): 2702–2715.
Schilbach, L., A.M. Wohlschlaeger, N.C. Kraemer, A. Newen, N.J. Shah, and G.R. Fink. 2006. Being with virtual others:Neural correlates of social interaction. *Neuropsychologia* 44(5): 718–730.
Schneider, D., Z.E. Nott, and P.E. Dux. 2014. Task instructions and implicit theory of mind. *Cognition* 133(1): 43–47.
Shimpi, P.M., N. Akhtar, and C. Moore. 2013. Toddlers imitative learning in interactive and observational contexts: the role of ageand familiarity of the model. *Journal of experimental child psychology* 116(2): 309–323.
Spaulding, S. 2010. Embodied cognition and mindreading. *Mind & Language* 25(1): 119–140.
Surian, L., S. Caldi, and D. Sperber. 2007. Attribution of beliefs by 13-month-oldinfants. *Psychological Science* 18(7): 580–586.
Surtees, A., and I. Apperly. 2012. Egocentrism and automatic perspective taking in children and adults. *Child Development* 83: 452–460.
Thierry, G., P. Athanasopoulos, A. Wiggett, B. Dering, and J.-R. Kuipers. 2009. Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences* 106(11): 4567–4570.
Winawer, J., N. Witthoft, M.C. Frank, L. Wu, A.R. Wade, and L. Boroditsky. 2007. Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences* 104(19): 7780–7785.
Woodward, A.L., J.A. Sommerville, S. Gerson, A.M. Henderson, and J. Buresh. 2009. The emergence of intention attribution ininfancy. *Psychology of learning and motivation* 51: 187–222.
Yoon, J.M., M.H. Johnson, and G. Csibra. 2008. Communication-induced memory biases inpreverbal infants. *Proceedings of the National Academy of Sciences* 105(36): 13690–13695.