Beyond 'Interaction': How to Understand Social Effects on Social Cognition

Julius Schönherr and Evan Westra

Abstract

In recent years, a number of philosophers and cognitive scientists have advocated for an 'interactive turn' in the methodology of social-cognition research: to become more ecologically valid, we must design experiments that are interactive, rather than merely observational. While the practical aim of improving ecological validity in the study of social cognition is laudable, we think that the notion of 'interaction' is not suitable for this task: as it is currently deployed in the social cognition literature, this notion leads to serious conceptual and methodological confusion. In this paper, we tackle this confusion on three fronts: 1) we revise the 'interactionist' definition of interaction; 2) we demonstrate a number of potential methodological confounds that arise in interactive experimental designs; and 3) we show that ersatz interactivity works just as well as the real thing. We conclude that the notion of 'interaction', as it is currently being deployed in this literature, obscures an accurate understanding of human social cognition.

1. Introduction

Consider the following example of a typical social interaction:

Ian and Mia

Mia enters a coffee shop and sees her best friend Ian sitting on the sofa. Ian doesn't notice her right away because he is stooped over his phone, closely examining the image of a woman on a dating website. Ian looks up and sees Mia, who smirks when she sees what he's been looking at. Ian blushes and quickly puts away his phone. 'It's not what you think', he says. 'I'm helping Sarah set up her profile'. Mia chuckles and asks Ian if he would like something from the barista. Ian asks for a green tea. On her way back from the counter, Mia trips, and spills both of their drinks all over her jeans. She looks around, and notices how everybody in the coffee shop is staring at her.

Now, contrast this with the following description of a standard false-belief task procedure, which is typical of social cognition research:

Standard False-Belief Task

Children see a toy figure of a boy and a sheet of paper with a backpack and a closet drawn on it. 'Here's Scott. Scott wants to find his mittens. His mittens might be in his backpack or they might be in the closet. Really, Scott's mittens are in his backpack. But Scott thinks his mittens are in the closet'. 'So, where will Scott look for his mittens? In his backpack or in the closet?' (the target question) 'Where are Scott's mittens really? In

his backpack or in the closet?' (the reality question). To be correct the child must answer the target question 'closet' and answer the reality question 'backpack'. (Wellman and Liu [2004])

Real-life social interactions like **Ian and Mia** are complex. They involve, among other things, belief ascriptions, gaze cues, emotional signals, gestures, relationships, and social conventions. Despite this complexity, the scientific study of such situations tends to rely on simplified, highly artificial paradigms like **Standard False-Belief Task**. Ostensibly, the kind of knowledge being tested in the false-belief task is also supposed to be the knowledge that Ian and Mia use in order to successfully navigate their social encounter—namely, their theory-of-mind. However, the difference between these two vignettes is hard to ignore. Of course, some might argue that for all their artificiality, we need tools like the false-belief task if we are ever to begin to make sense of how social cognition functions. This is a trade-off inherent to all experimental psychology: if we desire scientific rigor, we must sacrifice some ecological validity.

However, there are a number of theorists who think that experimental paradigms in social cognition research like **Standard False-Belief Task** sacrifice far too much (De Jaegher and Di Paolo [2007]; Gallagher and Hutto [2008]; Schilbach *et al.* [2013]). For instance, in this experiment the child is set apart from Scott. There is no possibility for the two to interact. There are no reciprocal gaze cues, no emotional signals, no gestures, and no relationships. The child is merely a passive observer. In **Ian and Mia**, on the other hand, both agents are interacting. They mutually respond to and transmit a wide range of social cues, which get interpreted in a context-sensitive fashion.

Interactionists conclude that, as a consequence of experimental oversimplification, traditional research on human social cognition has lost sight of the very phenomenon it set out to explain. What is needed, they propose, is an 'interactive turn' towards more 'second-personal' methods and theories that acknowledge the dynamic, interdependent aspects of ordinary social experiences. More specifically, according to interactionists, past research is problematic because it relies heavily upon observational experimental paradigms. Real social cognition, however, almost always takes place in interactive contexts. As a result, current theoretical and empirical paradigms are thought to be ill-suited to study the cognitive processes at work in real-life social activities.

One example of this kind of 'interactionist' approach to social cognition research is the double TV monitor paradigm (Murray and Trevarthen [1985]). In this experiment, 2-month old infants were shown a TV screen displaying a video of their mothers. In the 'interactive' condition, the video was live, while in the 'non-interactive' condition, the video showed a replay of their mother's actions. It was found that infants quickly disengaged when presented with the replay video, but were far more motivated to attend to the feed in the interactive condition.

Another example is the perceptual crossing paradigm (Auvray *et al.* [2009]; Auvray and Rohde [2012]). In this experiment, two players move an avatar along a one-dimensional strip using a computer mouse. When moving her own avatar along the strip, a player can cross paths with three objects: a static object, the other player's avatar, and the other player's avatar's shadow (i.e. an object copying the movements of the other player's avatar). Each agent receives the same sensory feedback upon crossing paths with any of these three

objects. Importantly, when one player's avatar crosses paths with another player's shadow only the player with the avatar receives feedback. If two players' avatars meet, both players receive sensory feedback. Interestingly, although the sensory feedback a player receives from crossing paths with any of the objects is identical, players nevertheless typically manage to 'find' one another (i.e. oscillate their avatars around each other).

The interactionist criticism, thus far, amounts to the claim that traditional research paradigms such as **Standard False-Belief Task** need to be supplemented by novel interactive paradigms; and this criticism is well-taken. However, some interactionists have taken their critique a step further, and argued that the socio-cognitive processes at work in interactive contexts are fundamentally distinct from those that operate in observational ones. For instance, Shaun Gallagher and Daniel Hutto have argued at length that mental state attribution really only occurs in observational scenarios; in social interactions, we rely upon a range of non-mentalistic processes, including gaze-following, social narratives, and emotional mirroring (Gallagher [2001], [2009], Hutto [2004], [2007]; Gallagher and Povinelli [2012]).

At times, interactionists in the enactivist tradition make the even more radical claim that social interactions can actually constitute social cognition. Interactions, it is argued, have emergent properties that cannot be reduced to contributions of individuals. When two autonomous agents act in such a manner that their actions are 'coupled' (i.e. causally interdependent), this can create a higher order 'dynamical system' with its own intrinsic properties. These systems, it is claimed, are the true loci of social cognition (De Jaegher *et al.* [2010]). We should, therefore, abandon the idea that social cognition can fully be explained in purely individualistic terms. Instead, according to the interactionists, social cognition researchers ought to focus their efforts on the intrinsic properties of interactive systems.

Other proponents of the interactive turn have de-emphasized the claim that social interactions are constitutive of social cognition. For instance, Gallotti and Frith have argued that interacting agents have 'novel routes to knowledge of other minds' that facilitate cooperation and team reasoning (Gallotti and Frith [2013], p. 162). This route to social knowledge is achieved by entering into the 'we-mode', a psychological state in which aspects of an interactive scene are represented via distinctively collective mental attitudes: believing-together, intending-together, desiring-together, etc. When agents enter the 'we-mode', they co-represent the action-possibilities available to their interactive partners, and use this information to make decisions that achieve collective ends. Andreas Roepstorff and colleagues have also proposed that social situations can be interactive to varying degrees; with increasing degrees of interactivity, they find corresponding effects upon processing speed (Tylén *et al.* [2012]), accurate collective decision-making (Bahrami *et al.* [2012]), and physiological and behavioural alignment (Fusaroli *et al.* [2016]).

While proponents of the interactive turn come in various flavors, they all endorse a central methodological claim: in order to promote ecological validity, experiments in social cognition need to become less observational and more interactive. In this paper, we will argue that this way of thinking is misguided. We are of course in favor of improving the ecological validity of social cognition research; however, we think that the notion of 'social interaction', as it is currently being deployed, is the wrong tool for the job. We argue that contrasting social cognition in interactive and non-interactive contexts is often uninformative, and prone to methodological confusion. This is because both the proximal

causes and underlying mechanisms that support naturalistic social cognition tend to straddle the interaction/observation dichotomy. In short, we believe that emphasizing 'interaction' is a red herring.

To show why this is the case, we will first turn our attention to the definition of 'interaction' that has become the standard in the interactionist literature. We will argue that this definition introduces concepts that needlessly complicate the target phenomena. In its place, we will offer a pared down, minimalist definition of 'interaction' that adequately captures the phenomena that interactionists are interested in.

Next, we will point out an obstacle to any cognitive scientist wishing to implement 'interactionist' experimental paradigms. This is that interactions are typically composed of many different social elements that are not themselves interactive. These concomitant social elements create a number of potential confounds for interactionist experiments, which social cognition researchers would do well to control for. To this end, we review four bodies of literature that illustrate the need for appropriate, non-interactive controls in interactionist paradigms: the 'Social Simon Effect', spontaneous perspective-taking, imitation, and conversational alignment.

Finally, we will argue that in many cases, so-called 'interactionist' paradigms have really featured ersatz interactions. We think this shows that it is not interaction as such that really makes a difference in social cognition research, but rather that individual participants believe themselves to be interacting. This contradicts the basic anti-individualist thrust of interactionism.

2. Defining 'Interaction'

We now turn to the issue of defining 'social interaction'. This turns out to be a delicate matter: while it is widely acknowledged that to develop an adequate theory of social cognition, we should be studying social interactions, there are ways of defining the term that largely presuppose a particular theory of social cognition. But if studying social interaction is supposed to provide evidence for these same theories, this ends up being circular. What is needed, rather, is a theory-neutral definition of social interaction that all interested parties can agree upon. This notion of interaction can then serve as a common point of departure for future debates. Therefore, our strategy in this section will be to start with the most prominent definition of social interaction in the extant literature, and then pare it down to a minimal, theory-neutral form.

The most influential definition of 'social interaction' comes from De Jaegher, Di Paolo, and Gallagher [2010]:

De Jaegher Interaction
Two or more autonomous agents co-regulating their coupling with the effect that their autonomy is not destroyed and their relational dynamics acquire an autonomy of their own. Examples: conversations, collaborative work, arguments, collective action, dancing and so on. (De Jaegher *et al.* [2010], p. 441)

An 'autonomous system' is further defined as a 'network of co-dependent, precarious processes able to sustain itself and define an identity as a self-determined system' (De Jaegher *et al.* [2010], p. 441). The set of autonomous systems, on this definition, includes most biological life-forms, from single-celled organisms to human beings, and also socially constructed entities, like corporations and nations. In the context of social cognition, the relevant class of autonomous systems is restricted to autonomous agents. 'Coupling' occurs when one autonomous system causally impacts the functioning of another. Coupling is said to be 'regulated' when this causal impact is in some way controlled by that system; and it is said to be 'co-regulated' when two or more autonomous systems are controlling how they causally impact one another. Genuine social interactions, on this view, occur when this co-regulated coupling results in the creation of a new autonomous system while still preserving the autonomy of the co-regulators. Lastly, this emerging interactive system is required to be temporally extended enough to take on 'autonomy' of its own.

Our first issue with this definition is related to the idea that genuine social interactions take on 'an autonomy their own'. As noted above, a definition of 'social interaction' should, where possible, be theory-neutral; it should not entail a particular social ontology. However, the ontology implied by the above phrase is highly controversial: namely, that interactions create new autonomous systems. These autonomous systems are then thought to form the proper objects of social cognition research: they literally constitute social cognition (De Jaegher *et al.* [2010]). But a number of authors have argued that this claim amounts to a confusion of constitution and causation (Herschbach [2012]; Carruthers [2015]). Given that this debate is still ongoing, it seems unnecessary to hardwire such a controversial metaphysical claim into a practical, theory-neutral definition. Therefore, we propose the following first initial revision to De Jaegher's *et al.*'s definition:

> **Interaction - First Revision**
> Two or more autonomous agents co-regulating their coupling with the effect that their autonomy is not destroyed.

Our second worry concerns the role that the concept of 'autonomy' plays in this definition. In De Jaegher and colleagues' definition, 'autonomy' is introduced as a technical notion according to which almost all biological life forms, not just human beings, can constitute autonomous systems (i.e. they can form self-sustaining and self-determining systems). Likewise, interactions between such autonomous systems don't necessarily have to involve human beings either: interactive systems would come into being whenever two cells cross paths in a petri dish, and whenever two countries engage in diplomatic negotiations. With such a broad scope, one might worry that this notion of social interaction is indeed too broad to be of any scientific utility. If the study of social cognition is to take an 'interactive turn', then interaction needs to be something that can be operationalized in a controlled, experimental setting.

Presumably, it is for these reasons that De Jaegher and colleagues narrow their definition to be specifically about autonomous 'agents'. However, in this case 'autonomy'—at least in the technical sense of the term—does not do any definitional work. This is because the set of agents is a proper subset of the set of autonomous systems. Therefore, the phrase 'autonomous agent' is not more informative than the term 'agent'.

Furthermore, given their technical notion of autonomy, it is unclear why cases of coercion should be discounted, as De Jaegher and colleagues maintain (De Jaegher and Di Paolo [2007], p. 495; De Jaegher et al. [2010], p. 443). In a case of armed robbery, for instance, it would seem that we have an instance of correlated mutual behaviour that is at least as complex as the case of two people having a conversation. Why, then, would this fail to create an interaction? According to De Jaegher and colleagues, the coercive nature of the mugger's actions would 'destroy the autonomy' of the victim. If the criteria for autonomy are so weak that bacteria in a petri dish can form an autonomous system, it is hard to see how it could be destroyed simply by demanding, 'Your money or your life!'. Even if the victim complies, it seems as though her status as an autonomous system in the sense being used here would be preserved.

Of course, there is a classic, Kantian sense in which the victim's autonomy in this situation is compromised—namely, her ability to act in accordance with a law of her own choosing. If interactionists were to adopt this notion of autonomy in their definition, they could avoid the charge of vacuity. However, we would then need to dramatically revise the range of cases that would count as social interactions. First, the subset of entities that possess autonomy in this strong sense will be much smaller than those that possess it in the weaker sense. Young children and animals, for instance, are unlikely to be autonomous in this sense. Drug addicts and persons with cognitive disabilities would also likely to fall below the threshold. Women in highly patriarchal countries with oppressive religious laws would also lack this kind of autonomy. Second, although human agents can be autonomous in this sense, it is unclear what it would mean for a co-regulated coupling to create an autonomous system. In short, it is not clear when—if ever—the conditions for interaction would obtain, given this notion of autonomy. Lastly, and most importantly, it is not at all clear why an obviously normative notion should play a role in cognitive science. The fact that a person cannot act in accordance with the law of her own choosing does not obviously bear on the cognitive mechanisms she brings to bear when encountering other agents.

These problems associated with the Kantian notion of autonomy also generalize to other normative theories of autonomy, which are generally unfit to constrain cognitive theories of interaction. To see this more clearly, consider the higher-order theory of autonomy defended by Michael Bratman (Bratman [2003], [2007]). According to Bratman, autonomous agents treat mere considerations to act as justifying reasons to act ([2007], p.178). Treating one's considerations in this way functions as a guide to resolve indecision and is, therefore, desirable. Autonomy, understood in this way, is a normative notion. Agents can fail to act autonomously, if they fail to have appropriate higher-order regard for their first-order motivations. Importantly, it is implausible that agents who fail to treat their considerations for action as justifying reasons cannot engage in mundane (but clear) forms of interaction (e.g. paying the cashier for the groceries I wish to buy). In short, normative theories of autonomy introduce constraints that are too restrictive to ground cognitive accounts of interaction.

Thus De Jaegher and colleagues' reliance on 'autonomy' in their definition faces a dilemma: given the original, more technical notion of autonomy, interactions are so ubiquitous and variable that they do not form a category of scientific interest. Given a more demanding, normative notion of autonomy, interactions become so rare that it is not clear whether they occur at all. Interactionists could address this issue by providing an alternative

account of 'autonomous systems' that is situated somewhere in between these two extremes. But until such an account is provided, the notion of 'autonomy' is not scientifically useful. Therefore, we propose a second revision to De Jaegher's definition:

Interaction - Second Revision
Co-regulated coupling between conscious human beings.

This revised definition does away with the notion that interactions must be performed by autonomous systems. But nothing serious is lost. We noted that once the relevant class of agents is specified, the further classification 'autonomous agents' is explanatorily inert. The revised definition makes explicit that, in the context of social cognition, the relevant class of agents are conscious human beings. To be sure, other types of organism may also engage in interactions, but this need not concern us.

Lastly, we propose a small addition to our definition: two agents or more co-regulate their coupling if the actors knowingly[1] affect each other's actions. This further specification is necessary to rule out cases in which agents affect each other's actions by mere accident. Consider the case in which you swipe the foliage from your lawn into my lawn. I, thinking that a sudden gust of wind is responsible, swipe it back into your lawn. You, having the same thought, swipe it back into my lawn. We keep doing this until the end of August, when the foliage finally decays. Although we're affecting each other, we are, intuitively, not interacting. Moreover, our behaviour is uninteresting from the perspective of social psychology. Lastly, the addition of knowingly is preferable to the addition intentionally, because it does not exclude cases in which several agents affect each other's actions by mere foresight[2].

In summary: after a few clarificatory modifications of De Jaegher and colleagues' account of interaction, we are left with the following definition.

Minimal Social Interaction
When two or more conscious human beings mutually and knowingly affect one another's actions, they are engaged in a social interaction.

This minimalist definition fits nicely with paradigmatic examples of social interaction: conversation, dancing, cooking a meal together, playing tennis, etc. It also does not, however, eliminate cases of coercion and manipulation, such as the mugger scenario, or even actively violent encounters, such as fistfights. But it is not clear why these cases should be eliminated: surely, not all social interactions are pleasant and cooperative. While we may morally disapprove of these actions, this does not make them any less interactive.

This minimalist definition also fits nicely with key examples of interactionist experiments. In the Double TV-monitor paradigm, for instance, the live-feed condition makes it so that infants and their mothers are able to mutually respond to one another's actions; when the

---

[1] Note that, for our purposes, knowingly should be given a deflationary reading that is common in psychology (Dienes and Perner [1999]; Nagel [2013]). Knowing X, in this sense, means 'being aware of X and being sensitive to X when acting'. For instance, for Dienes and Perner (1999) mere perceptual awareness is sufficient for knowledge. What is more, having knowledge does not require recognizing that one has knowledge; i.e. it does not presuppose the concept KNOWLEDGE. Lastly, interacting knowingly does not presuppose the concept of INTERACTION; rather, it merely requires being aware of the constituents of interaction (e.g. the other person's voice and actions).

[2] Think, for instance, of a case in which you merely intend to get the foliage off your lawn, but you also foresee that I'll be mad when I find the foliage on my lawn. However, you don't intend to make me mad; you merely foresee that this will happen.

recording of the mother's expressions are played back for the child, this is no longer possible. In the perceptual-crossing study, participants are able to locate one another's sensors on the one-dimensional strip because they are able to mutually respond to one another, whereas the 'shadow' and the fixed object cannot.

According to the minimal approach, paradigms like the standard false-belief task would not count as interactive. This is because the actions of the character in the vignette do not affect the child's actions, and the child's actions do not affect those of the character in the vignette. The child merely observes the events taking place in the vignette, and then makes a prediction about them. There is no opportunity for a reciprocal exchange of information between the child and the character, nor any possibility for mutuality. It is decidedly non-interactive.

With this definition in hand, we are now in a position to defend our main point: if we want to improve the ecological validity of social cognition research, we should not frame this effort in terms of a distinction between interactive and observational scenarios.

3. The Constituents of Interaction

Proponents of an 'interactive turn' in social cognition research claim that in order to learn more about the nature of social cognition, we need to create more interactive experimental designs, and get away from purely observational paradigms. There is nothing wrong with designing interactive paradigms; however, it's not clear how much we really learn when we try to directly compare interactive and non-interactive contexts. This is because social interactions typically involve many different elements that are not themselves interactive.

To illustrate, take a prototypical interaction: a conversation with a colleague by the drinking fountain. Such an encounter would involve the physical co-presence of two individuals; however, this by itself would not make it an interaction. Likewise, the two speakers might possess mutual background knowledge about one another, including beliefs about each other's occupation, political views, short- and long-term goals, and so on. But this too does not make the encounter an interactive one. The conversation also involves the use of language. But even this, all by itself, fails to make the context interactive: one could easily imagine a person speaking aloud to herself, while another person ignores her. None of these elements, by themselves, it seems, are enough to make an encounter interactive. But all the same, they seem to be very important elements of the context, from a cognitive perspective.

Social interactions like this one seem to be complex events, composed of many elements that contribute to its interactive nature, and yet are not themselves interactive. All of these elements—physical co-presence, background knowledge, the use of language—often co-occur in social interactions, but are neither necessary nor sufficient for an interaction to occur. But, as we shall see in this section, they still have considerable effects on social cognition. As such, it is unclear whether 'interactive' effects on social cognition are driven by interaction as such, or by one of its component elements. In this section, we use several distinct bodies of evidence to argue that simply contrasting interactive and non-interactive

scenarios is not informative. This, we claim, reveals a key oversight in the interactionist approach.

3.1. The Social Simon Effect (Sebanz et al. [2003])

In a typical 'Simon' task, subjects carry out responses using their left and right hands to stimuli appearing on the left and right sides of a screen; typically, subjects are faster to respond to stimuli appearing on the side congruent with the response (i.e. left side of the screen with left hand response), and slower to respond to items appearing on the incongruent side (i.e. left side of the screen with right hand response) (Craft and Simon [1970]). Natalie Sebanz and colleagues modified this task so that it involved two subjects participating in parallel to one another, each responsible for responding with either the left or right hand; thus, subjects only had to respond in a Go/No-Go fashion depending on what they saw on the screen, regardless of which side the stimuli appeared on (Sebanz et al. [2003]). Importantly, their performance in no way depended upon what the other agent did—all they ever had to do was pay attention to their own screen and respond accordingly. Thus, there was nothing interactive about the task.

When subjects performed this task alone in a control condition, there was no spatial congruency effect—they were equally quick to respond to items on either side of the screen. But in the social condition, there was a spatial congruency effect: subjects were slower to respond to items on the side opposite their response hand (and on the same side as the other participant's response hand). In effect, the presence of another agent altered the way they represented their environment, such that it included both their own action affordances, and those of the other agent. Even when seated side-by-side with another agent completing totally independent tasks, their sheer presence affects how we represent and respond to the environment.

Since Sebanz and colleagues discovered the Social Simon Effect, a number of other experiments using similar paradigms have replicated and extended this finding. Using variants of the Social Simon paradigm, Guagnano and colleagues found that the Social Simon Effect dissipated with increased spatial separation between the two agents (i.e. within or beyond arm's length) (Guagnano *et al.* [2010]); Vlainic and colleagues found that the effect persisted even when subjects had no online perceptual feedback from the other participant, demonstrating that simply knowing that another agent is completing a similar task is enough to alter how one represents one's own action space (Vlainic *et al.* [2010]). Freundlieb and colleagues showed that when another agent was co-present but inactive, or co-present but completing a task of which the subject was ignorant, the effect dissipates (Freundlieb *et al.* [2015]).[3]

Thus, simply knowing that another agent is acting nearby is enough to alter the way that we respond to our environment, even when no interaction—even in the minimal sense—is taking place. Given that most interactive experimental designs include the co-presence of active agents, it may be that co-presence effects—which are not, in fact, the products of

---

[3] Guagnano *et al.* ([2010]) interpret their results as showing that the Social Simon Effect is due to participants representing their own action space, not the action affordances of those around them. But this claim in undermined by the results of Vlainic *et al.* ([2010]) and Freundlieb *et al.* ([2015]), which show that knowledge of another agent's action is key to generating the spatial congruency effect.

interaction—also occur in those tasks. This creates a methodological confound for proponents of the 'interactive turn' in experimental design: how are we to know whether purported interaction effects are genuine, or simply the product of the co-presence of other active agents?

3.2. Level-2 perspective-taking

Physical co-presence also seems to have an effect upon whether or not we spontaneously engage in certain forms of perspective-taking, the representation of what another agent can see. Psychologists typically distinguish between two 'levels' of perspective-taking (Masangkay *et al.* [1974]; Flavell *et al.* [1981]): Level-1 perspective-taking means representing whether or not a particular object is in the visual field of an agent, and is sensitive to external, environmental factors like line-of-sight and occlusion (Michelon and Zacks [2006]). Level-2 perspective-taking further involves the ability to represent how an object appears to another agent; for instance, the numeral '6' might, from one angle, appear to represent the number six, and from another angle, appear to represent the number nine; sensitivity to these differences requires an understanding of the aspectual nature of perception (Surtees et al. [2012], [2016]). Until recently, our best evidence suggested that while Level-1 perspective-taking is automatic and effortless, Level-2 perspective-taking is effortful and requires top-down, intentional control (Qureshi *et al.* [2010]; Samson *et al.* [2010]; Surtees e*t al.* [2012]). However, the relevant perspective being taken in these tasks was always that of a non-descript, computer generated avatar. But when the avatar is replaced with a live agent, we see a very different effect (Elekes *et al.* [2016]).

In this experiment, subjects sat in front of a monitor lying flat in front of them, and had to verify whether or not the numeral on the screen matched a number they heard in an audio recording. In the Individual condition, subjects completed this task alone; in the Joint condition, subjects sat opposite another participant who was either also completing a number-verification task (i.e. the perspective-dependent task), or a different task in which they had to say whether the colour of the numeral on the screen was the same as one they'd seen just previously (i.e. the perspective-independent task). Participants in the joint condition always knew which task the person opposite them was completing. Importantly, all subjects had to do was complete their own task—the actions of the other agent were always irrelevant. Thus, the task was not interactive (given our definition).

Elekes and colleagues found that subjects in the Joint condition were slower and made more errors than in the Individual condition, but only when both completed the perspective-dependent task and the numerals of the screen were such that their values differed on the basis of perspective (i.e. 2, 5, 6 and 9); for numerals whose values appeared to be the same regardless of which side of the table the participant was at (i.e. 0 and 8), there was no difference between the Individual and Joint conditions. In effect, subjects were only slower when 1) they had a live partner, 2) they believed that their partner had a similar goal, and 3) the partner's response would diverge from their own on the basis of their Level-2 perspective. In other words, when subjects knew that the person across the table from them was viewing the numeral on the screen as a number, they spontaneously maintained a representation of what he or she saw, and this representation then interfered with their own performance.

Thus, in this task, the mere co-presence of an active agent was not sufficient to prompt Level-2 perspective-taking, but the combination of co-presence and the knowledge that this agent had a goal similar to their own did. These results complement those of the Social Simon Task: when another agent is co-present, active, and has a goal similar to our own, we spontaneously represent both how the environment appears to them, and the kinds of actions that are available to them in that environment.

In interactive scenarios, of course, we are usually aware of the physical presence of other agents and their goals. Thus, we might expect that in those scenarios, we would also represent the affordances of the environment differently, or spontaneously adopt our partner's visual perspective. Upon observing all of these levels of socio-cognitive processing layered on top of one another, it is tempting to hypothesize that social interactions are irreducibly complex, and possess emergent properties. However, many of the constituents of this interaction are indeed isolable, and we can study the effects of these constituents individually. Moreover, we know that these social effects on social cognition are not inherently interactive, because we can also observe them in non-interactive scenarios. This is, we think, the central problem with the 'interactive turn': by focusing on interaction as a global property of social-cognitive scenarios, we miss out on a wealth of local, fine-grained information that may be present in non-interactive contexts.

A proponent of the 'interactive turn' could object that the cases we've described here are in fact best understood as effects of 'we-mode' cognition (Gallotti and Frith [2013]). For even though subjects in the Social Simon tasks and the perspective-taking task are not yet engaging in an interaction, they may be cognitively preparing for an interaction. The sheer proximity of their partners and the similarity of their tasks, the interactionist might argue, creates the sense that they are about to interact with one another, and this leads them to become more sensitive to their partner's perspective and action possibilities. Alternatively, these contexts might be said to create the illusion of interaction, where in fact there is none. Either way, the objection might go, these effects really only make sense in an interactionist framework.

We think that this objection makes an important point, but also a crucial concession. It may well be true that the cognitive processing that takes place in these near-interactive contexts have the function of supporting interaction. However, the fact remains that their presence was revealed in a non-interactive context, and that interaction is not necessary for eliciting them. Rather, the non-interactive task-design was a crucial part of discovering these processes. Thus, even if interaction might be a part of the explanation of why these effects are present, it was crucial that interaction was not a part of the task that revealed them.

In sum, it is important to identify the various sub-components of interaction, and not to mistake the effects of these sub-components for effects of the interaction itself. In practice, this will mean employing experimental paradigms that are explicitly non-interactive.

3.3. Interaction effects on infant learning

One line of research that seems to emphasize the importance of interactive methods is the literature on 'natural pedagogy' (Gergely and Csibra [2005]; Csibra and Gergely [2006],

[2009b]). According to this view, when an infant is addressed with certain ostensive signals (e.g. eyebrow-raising, eye contact, infant-directed speech), children spontaneously adopt a specialized learning stance. This learning stance prepares children to attend to certain kinds of information, such as facts about the identity and category-membership (Csibra and Gergely [2009a]). The pedagogical stance is also said to facilitate imitative learning.

The natural pedagogy hypothesis is not an explicitly interactionist proposal. However, it does seem to buy into the central methodological prescription of interactionism: there are certain forms of cognition that can only be studied in interactive contexts. Experiments in this tradition also frequently use observational controls to demonstrate the effects of pedagogical learning. For example, Yoon and colleagues found that 9-month olds tended to encode information about the location of an object in a non-interactive context, but instead encoded information about the object's identity in an interactive context with pedagogical cues (i.e. where an experimenter engaged in infant-directed speech and eye-contact) *(Yoon et al.* [2008]). The authors suggest that this is because interactive, pedagogical contexts prompt children to pay special attention to generic information. Likewise, in a study with 14- to 16-month-olds, Brugger and colleagues found that infants were more likely to imitate novel actions more in interactive, pedagogical contexts than in observational contexts *(Brugger et al.* [2007]). Based on these contrastive observational-versus-interactive designs, proponents of the natural pedagogy hypothesis argue that pedagogical interactions trigger specialized learning mechanisms that are not active in observational contexts.

The natural pedagogy hypothesis, however, remains controversial if cast as a theory specifically about interaction. To see why, note that in Brugger *et al.* ([2007]) and Yoon *et al.* ([2008]), the non-interactive condition was both non-communicative (i.e. the action was demonstrated by a solitary person) and observational (i.e. the child was not addressed through ostensive cues). Communicative contexts, however, are not necessarily interactive: one can observe communication between third parties without actively participating in it. Hence, these experiments leave open the possibility that the same learning effects attributed to pedagogical interactions might also occur in observational but communicative contexts.

Once the relevant distinctions are introduced, the importance of interaction in imitative learning becomes much less obvious. For instance, (Matheson et al. [2013]) conducted a study in which 18-month-olds and 24-month-olds imitated novel actions (e.g. ringing a doorbell using one's forehead) in (a) an interactive condition in which the experimenter addressed the infant using typical ostensive cues, (b) an observational and non-communicative condition in which the infant watched the experimenter perform the novel action all by herself, and (c) an observational-communicative condition, in which the infant watched the experimenter perform the novel actions while demonstrating them to another person. They found that 18-month-olds imitated more in the interactive condition than in the observational–non-communicative condition, but not significantly more than in the observational–communicative condition. In other words, it was the communicative dimension of the interactive condition that seemed to have improved imitation, rather than

interaction as such. In 24-month-olds, meanwhile, there were no differences across all three conditions.[4]

Shimpi and colleagues achieved a similar result while also manipulating the child's familiarity with the imitative model (e.g. whether the model was a family member, a complete stranger, or a stranger with whom the child had briefly interacted[5] before the task began) (Shimpi *et al.* [2013]). Interestingly, children in the observational-communicative condition imitated consistently regardless of whether they were familiar with the model; in contrast, children in the interactive condition imitated far less with unfamiliar models than familiar models. Thus, while children were quite adept at learning imitate complete strangers in observational–communicative contexts, some familiarity with the model was a prerequisite for imitative learning in interactive contexts.

On the one hand, these experiments do suggest that interaction can facilitate imitative learning in infants. However, these effects are not particularly pronounced: in Matheson *et al.* (2013) imitative learning in the older children was the same for all three conditions; in Shimpi *et al.* ([2013]) observational learning in communicative contexts was robust; and interactive learning was crucially dependent on the familiarity of the actor. The importance of interaction in imitative learning thus appears to be overstated. Similar observational-communicative controls have yet to be carried out for other forms of learning described by the natural pedagogy hypothesis (e.g. generic learning), and we cannot say for certain whether observational learning will be equally robust in that domain. However, we think there is good reason to find out.

3.4. Conversational alignment

We've noted that there are several social factors that are present in many social interactions, that have noticeable effects on social cognition, and that might be mistaken for interaction effects, but which are in fact non-interactive. However, an interactionist might object, even if these factors are present in non-interactive scenarios, they may still have unique effects in the context of a social interaction. Take, for instance, our paradigmatic example of a social interaction: conversation. We have pointed out that language use, by itself, is not inherently interactive. But, the interactionist might insist, language works much differently when studied as monologue than when it is studied as dialogue.

This is the central point behind the 'interactive alignment' research program of Martin Pickering and Simon Garrod, which has focused on the nature of speech production and comprehension during naturalistic dialogues (Garrod and Pickering [2004], [2009], Pickering and Garrod [2004], [2013]). Explicit in this research program is a critique of psycholinguistic theories based on the study of comprehension and production of speech in non-interactive

---

[4] Interestingly, emulation was significantly higher in the solitary–non-communicative condition than in the interactive condition. (An actor's action is said to be emulated by an agent, if the actor's goal is copied by the action. An action is said to be imitated, if the agent copies the actor's exact action sequence.)

[5] Familiarity was established through a 10-minute warm-up period in which the experimenter played a sorting game with the child.

contexts (i.e. monologue). The most natural and basic form of language use, they claim, is dialogue; to develop a full understanding of the mechanisms of language, we need to study it in this form.

Central to Pickering and Garrod's positive account is the observation that speakers in a dialogue will tend to converge upon matching representations at the lexical, semantic, and syntactic levels—a phenomenon the authors call 'conversational alignment'. For instance, syntactic alignment refers to the spontaneous tendency of a speaker to use a particular syntactic construction when that same construction has just been used by an interlocutor (e.g. the cowboy gives the pirate a banana versus the pirate gives the banana to the pirate (Pickering and Branigan [1999]; Branigan *et al.* [2000], [2007])). In dialogue, this alignment of representations is said to take place at multiple levels simultaneously, with alignment at one level facilitating alignment at other levels through the co-activation of multi-level associative networks. As a result of this alignment process, participants in a dialogue achieve a high level of communicative fluency. This enables them to rapidly recover meaning from each other's utterances, even when these utterances are otherwise fragmentary, overlapping, and entirely ungrammatical. Other researchers have also extended the study of alignment in dialogue beyond the coordination of linguistic representations, and found evidence for analogous forms of synchronization in eye movements (Dale *et al.* [2011]) and heart-rate (Fusaroli *et al.* [2016]).

We agree with the general project of studying dialogue in naturalistic circumstances. However, we argue that much of Pickering and Garrod's own account of the mechanisms supporting conversational alignment depends upon evidence from individualistic paradigms. Moreover, while there are some differences in the magnitude of the relevant effects when these are measured in interactive contexts, these differences are readily explained in terms of other non-interactive mechanisms, such as increased attention. Finally, even where we do find uniquely interactive alignment effects, individualistic mechanisms still play an important role in their explanation.

For instance, Garrod and Pickering have suggested that alignment between speakers and listeners is a product of representational processes that are shared between the comprehension and production systems. Thus, when a listener hears an utterance of a sentence with a certain syntactic form or lexical item, those representations are primed for use in speech production. However, much of the evidence that Pickering and Garrod present for this mechanistic hypothesis is derived from non-interactive tasks (i.e. 'monologue'). For instance, the 'structural persistence' or priming of syntactic forms from comprehension to production has been established in numerous individualistic experimental paradigms, which Pickering and Garrod cite as evidence (Bock [1986]; Bock *et al.* [2007]). Pickering and Garrod (2013) also suggest that the shared representational processes in comprehension and production are the product of forward modelling mechanisms for action-planning (Davidson and Wolpert [2004]; Tourville *et al.* [2008]) that have been repurposed for the covert imitation and prediction of observed actions (i.e. mirror neurons (Gallese *et al.* [1996]; Umiltà *et al.* [2001])). But again, the evidence for such mechanisms is

drawn from paradigms that are entirely individualistic (Watkins *et al.* [2003]; Pulvermüller *et al.* [2006]; Ito *et al.* [2009]; Möttönen and Watkins [2009]; Adank and Devlin [2010]). Far from being irrelevant to our understanding of language, it seems that our understanding of interaction effects in language actually depends upon evidence gathered in non-interactive paradigms.

While the mechanisms underlying various alignment phenomena are present in non-interactive contexts, the case could be made that these mechanisms behave differently in social interactions. Branigan and colleagues ([2007]), for example, developed an interactive paradigm in which they were able to compare the rates of syntactic priming in participants in a conversational interaction with those in individuals who were merely side-participants. While they found syntactic priming effects in both groups, these effects were significantly stronger when a speaker had just been addressed than when he or she was merely listening to other individuals speak; but, as Branigan and colleagues themselves note, this effect is likely due to the fact that current addressee's were attending to the speaker more carefully than side-participants. Increased attention, of course, is not a uniquely interactive phenomenon. This suggests that while alignment does increase in the context of conversational interactions, alignment is nevertheless explained by a host of mechanisms that do not operate only in interactive contexts.

There are some aspects of conversational alignment that are, in fact, uniquely interactive. For instance, Garrod and Pickering ([2009]) describe how participants in a dialogue also coordinate upon the timing of their utterances, which tends to yield fairly precise patterns of turn-taking (Ten Bosch *et al.* [2004]; Levinson [2016]). This phenomenon truly has no non-interactive equivalent, since turn-taking is by definition impossible in a monologue. We happily concede that this might be a case where an interactive context is necessary to truly grasp the nature of the phenomenon. However, Garrod and Pickering's explanation for our capacity for precise turn-taking in conversation invokes precisely the same covert imitation and priming mechanisms that explain other aspects of alignment. Thus, even if our knowledge of this phenomenon depends upon interactive experimental designs, we owe our understanding of it to individualistic research.

Thus, while dialogue is often cited as a paradigm case of an irreducibly interactive process, we would argue that conversational alignment arises from mechanisms that are not inherently interactive. In some cases, we do see these mechanisms operating differently in the context of interaction. In the case of turn-taking, we seem to have an instance of a genuine interaction effect. But other properties of dialogue, such as syntactic alignment, are also present in monologue; indeed, our very understanding of this aspect of dialogue is due to its study in non-interactive contexts.

4. How Much Does 'Real' Interaction Matter?

It is sometimes suggested that interaction dynamics cannot be explained if we only look at the sum of the interactors' individual contributions to the encounter.[6] We don't wish to take a final stand on these issues in this paper. In this section, we'd simply like to point out that most of the interactionists' own experiments seem to tacitly presuppose an individualist framework.

In a series of experiments, Schilbach and colleagues have investigated interaction-specific neural activation patterns of action-control (Schilbach *et al.* [2011]), joint attention (Schilbach *et al.* [2010a]), and mutual gaze (Schilbach *et al.* [2006a]). In most of these experiments a subject is placed in an fMRI scanner engaging in some kind of interaction with a virtual character. Roughly, these experiments indicate that cues associated with interaction such as self-directed gaze are associated with differential neural activation in the medial prefrontal cortex, which is a region thought to be crucially implicated in social cognition (Van Overwalle [2009]). For instance, Schilbach finds differentially increased neural activation in the medial prefrontal cortex for (a.) direct (vs. other-directed) gaze (Schilbach *et al.* [2006b]), and for following (vs. leading) someone's gaze (Schilbach *et al.* [2010b]). To account for the interactive element, all participants are made to believe that the virtual character is controlled by a real person with whom the interaction will subsequently take place. This belief, however, was false: the virtual character was entirely preprogrammed to establish conditions of a controlled experiment. As a result, participants are not actually interacting. In terms of experimental design, this is fine; but what these experiments tacitly presuppose is that a subject's individual representation of a situation as interactive is sufficient to gain crucial insights in the cognitive significance of interaction.

One notable exception departing from the virtual-character paradigm is a study conducted by Cavallo and colleagues ([2015]). In this study, subjects established eye contact with a collaborator who was situated behind the fMRI scanner. The collaborator was visible to the participant via a mirror placed inside the scanner. In the experiment, either both subjects looked at each other (i.e. mutual gaze) or one of them looked away (i.e. averted gaze). In the control conditions participants either looked at their own eyes in a mirror reflection, or they looked at an image of the collaborator. Cavallo and colleagues found that mutual gaze differentially activates the anterior portions of the medial prefrontal cortex (mPFC).

As indicated above, Schilbach and colleagues found similar patterns of activation, even though they relied upon paradigms that used virtual characters (Schilbach *et al.* [2006b], [2010b]). Comparing these experiments, it seems that real interaction does not seem to have made a crucial difference to activation in the mPFC, which was the main finding in the

---

[6] For instance, De Jaegher *et al.* argue that 'interactive processes [...] complement and even replace individual mechanisms' (De Jaegher *et al.* [2010], p. 441). At the heart of this proposal is the idea that partitioning social cognitive processes into the cognitive mechanisms implemented by individual brains is unwarranted. Rather, it is the interaction between brains that should be considered explanatorily basic.

mutual gaze condition. Furthermore, Cavallo and colleagues found that neural activation was independent of whether subjects actually established eye contact or whether subjects merely knew that the collaborator was looking at them. Hence, it was the 'mere belief of being seen' (Cavallo et al. [2015], p. 67) which accounted for the distinct pattern of neural activation; actual interaction seemed irrelevant. Importantly, while experiments by Schilbach *et al.* support the idea that even simulated interaction leads to activation in the mPFC, the study by Cavallo *et al.* provides direct comparative evidence for the claim that real interaction is not crucial for the relevant neural activation patterns to occur. Lastly, while Schilbach also reports increased activity in the amygdala, Cavallo finds no such activity.[7] And even if differential activation in the amygdala were to indicate a difference between virtual and real interactions, the absence of such activity in a real interactive conditions is rather bad news for the interactionists, who have pointed out that emotional engagement is a crucial cognitive element in social interactions (Reddy [2008]; Schilbach *et al.* [2013]).

Together these observations suggest that, at least in gaze paradigms, it is more significant whether a subject believes that she engaged in an interaction; and not so much whether she is actually engaged in an interaction.

5. Conclusion

Our aim in this paper has been to draw attention to the various conceptual and methodological confusions that arise when we over-emphasize the notion of interaction in social cognition research. First, we argued that De Jaegher and colleagues' prominent definition of interaction diverged significantly from the intuitive consensus, and also seems to equivocate on the notion of autonomy. Second, we illustrated how interactive paradigms potentially confound genuine interaction effects with the effects of factors that merely co-occur with interaction. Finally, we showed that genuine interactions are not needed to study the effects of interaction on cognition: the mere representation of interactivity will often do just as well. Genuine interactivity, although often the distal cause of such representations, do not play a special role in explaining these effects.

However, our goal in this paper is not completely negative, and we are not wholly opposed to interactive experimental designs; rather, we advocate for a complementary, multi-method approach that includes both interactive and non-interactive methods. However, when interactive designs are used, we advise that researchers remain cautious in their interpretations, and that they implement appropriate controls before attributing the effects they discover to interaction as such. We hope that by drawing attention to the various confounds and confusions that arise in interactive experimental designs, we have clarified the significance of interaction in social cognition research. With this added clarity, we hope, researchers will now be better positioned to pursue the goal of making experimental

---

[7] Notably, involvement of the amygdala has been inconsistent throughout an array of studies investigating mutual gaze. For instance, while a number of authors (Kawashima *et al.* [1999]; Wicker *et al.* [2003]; Sato et al. [2004]; Schilbach *et al.* [2006a]) have found activation in the amygdala during mutual gaze, several others have not (Calder *et al.* [2002]; Pageler *et al.* [2003]).

paradigms in social cognition research more ecologically valid. With this end in mind, we have three general suggestions for future research:

1. Interaction is complicated, but defining it doesn't have to be:

   While the philosophical debate surrounding the ontology of social interaction is still ongoing, this debate need not impinge upon practical applications of the concept of interaction in research contexts. The notion of autonomy, in particular, serves merely to obscure, rather than to clarify, the meaning of 'social interaction'. In lieu of the one provided by De Jaegher and colleagues, we have offered our own definition that captures the intuitive notion of social interaction with minimal conceptual baggage.

2. Interaction effects versus social effects on social cognition:

   Ordinary social interactions are complex events, which tend to involve a cluster of social elements that are not themselves interactive. This makes it difficult to study the effects of interaction as such, because we must distinguish the effects of interaction from concomitant social factors. Researchers interested in improving upon the ecological validity of social cognition paradigms must recognize these factors could potentially dissociate from interaction, and ought to be investigated in their own right.

3. Real versus represented interaction:

   Many of the purported effects of interaction on social cognition can also be found in pseudo-interactive paradigms. This shows that paradigms manipulating beliefs about interaction can be just as informative as the paradigms that involve genuine interaction. Once this individualist insight into the 'interactionist turn' is taken on board, it opens up practical possibilities for social cognition research by making the problem of social interaction more empirically tractable.

Julius Schönherr
Department of Philosophy
University of Maryland
4300 Chapel Drive, College Park, MD 20742
schoenherrjulius@gmail.com

Evan Westra
Department of Philosophy
University of Maryland

4300 Chapel Drive, College Park, MD 20742
ewestra@umd.edu

References:

Adank, P., and Devlin, J.T. [2010]: 'On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech', *NeuroImage*, **49**, pp. 1124–32.

Auvray, M., Lenay, C., and Stewart, J. [2009]: 'Perceptual interactions in a minimalist virtual environment', *New ideas in psychology*, **27**, pp. 32–47.

Auvray, M., and Rohde, M. [2012]: 'Perceptual crossing: the simplest online paradigm', *Frontiers in human neuroscience*, **6**.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., and Frith, C. [2012]: 'Together, slowly but surely: The role of social interaction and feedback on the build-up of benefit in collective decision-making.', *Journal of Experimental Psychology: Human Perception and Performance*, **38**, pp. 3–8.

Bock, K. [1986]: 'Syntactic persistence in language production', *Cognitive Psychology*, **18**, pp. 355–87.

Bock, K., Dell, G.S., Chang, F., and Onishi, K.H. [2007]: 'Persistent structural priming from language comprehension to language production', *Cognition*, **104**, pp. 437–58.

Branigan, H.P., Pickering, M.J., and Cleland, A.A. [2000]: 'Syntactic co-ordination in dialogue', *Cognition*, **75**, pp. 13–25.

Branigan, H.P., Pickering, M.J., McLean, J.F., and Cleland, A.A. [2007]: 'Syntactic alignment and participant role in dialogue', *Cognition*, **104**, pp. 163–97.

Bratman, M.E. [2003]: 'Autonomy and hierarchy', *Social Philosophy and Policy*, **20**, pp. 156–76.

Bratman, M.E. [2007]: 'Structures of agency: Essays', New York, NY: Oxford University Press.

Brugger, A., Lariviere, L.A., Mumme, D.L., and Bushnell, E.W. [2007]: 'Doing the Right Thing: Infants' Selection of Actions to Imitate From Observed Event Sequences', *Child Development*, **78**, pp. 806–24.

Calder, A.J., Lawrence, A.D., Keane, J., Scott, S.K., Owen, A.M., Christoffels, I., and Young, A.W. [2002]: 'Reading the mind from eye gaze', *Neuropsychologia*, **40**, pp. 1129–38.

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

Cavallo, A., Lungu, O., Becchio, C., Ansuini, C., Rustichini, A., and Fadiga, L. [2015]: 'When gaze opens the channel for communication: Integrative role of IFG and MPFC', *NeuroImage*, **119**, pp. 63–9.

Craft, J.L., and Simon, J.R. [1970]: 'Processing symbolic information from a visual display: Interference from an irrelevant directional cue.', *Journal of Experimental Psychology*, **83**, pp. 415–20.

Csibra, G., and Gergely, G. [2006]: 'Social learning and social cognition: The case for pedagogy', *Processes of change in brain and cognitive development. Attention and performance XXI*, pp. 249–74.

——— [2009a]: 'Natural pedagogy', *Trends in cognitive sciences*, **13**, pp. 148–53.

——— [2009b]: 'Natural pedagogy.', *Trends in cognitive sciences*, **13**, pp. 148–53.

Dale, R., Kirkham, N.Z., and Richardson, D.C. [2011]: 'The dynamics of reference and shared visual attention', *Frontiers in Psychology*, **2**, pp. 1–11.

Davidson, P.R., and Wolpert, D.M. [2004]: 'Internal models underlying grasp can be additively combined', *Experimental Brain Research*, **155**, pp. 334–40.

De Jaegher, H., and Di Paolo, E. [2007]: 'Participatory sense-making', *Phenomenology and the cognitive sciences*, **6**, pp. 485–507.

De Jaegher, H., Di Paolo, E., and Gallagher, S. [2010]: 'Can social interaction constitute social cognition?', *Trends in cognitive sciences*, **14**, pp. 441–7.

Dienes, Z., and Perner, J. [1999]: 'A theory of implicit and explicit knowledge', *Behavioral and Brain Sciences*, **22**, pp. 735–808.

Elekes, F., Varga, M., and Király, I. [2016]: 'Evidence for spontaneous level-2 perspective taking in adults', *Consciousness and Cognition*, **41**, pp. 93–103.

Flavell, J.H., Everett, B. a., Croft, K., and Flavell, E.R. [1981]: 'Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction.', *Developmental Psychology*, **17**, pp. 99–103.

Freundlieb, M., Kovács, Á.M., and Sebanz, N. [2015]: 'When Do Humans Spontaneously Adopt Another's Visuospatial Perspective?', 41.

Fusaroli, R., Bjørndahl, J.S., Roepstorff, A., and Tylén, K. [2016]: 'A heart for interaction: Shared physiological dynamics and behavioral coordination in a collective, creative construction task.', *Journal of Experimental Psychology: Human Perception and Performance*, **42**, pp. 1297–310.

Gallagher, S. [2001]: 'The practice of mind. Theory, simulation or primary interaction?', *Journal of Consciousness Studies*, **8**, pp. 83–108.

Gallagher, S., and Hutto, D. [2008]: 'Understanding others through primary interaction and narrative practice', *The shared mind: Perspectives on intersubjectivity*, Philadelphia, PA: John Benjamins Publishing Company, pp. 17–38.

Gallagher, S. [2009]: 'Two problems of intersubjectivity', *Journal of Consciousness Studies*, **16**, pp. 289–308.

Gallagher, S., and Povinelli, D.J. [2012]: 'Enactive and Behavioral Abstraction Accounts of Social Understanding in Chimpanzees, Infants, and Adults', *Review of Philosophy and Psychology*, **3**, pp. 145–69.

Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. [1996]: 'Action recognition in the premotor cortex.', *Brain*, **119**, pp. 593–609.

Gallotti, M., and Frith, C.D. [2013]: 'Social cognition in the we-mode', *Trends in Cognitive Sciences*, **17**, pp. 160–5.

Garrod, S., and Pickering, M.J. [2004]: 'Why is conversation so easy?', *Trends in Cognitive Sciences*, **8**, pp. 8–11.

————— [2009]: 'Joint Action, Interactive Alignment, and Dialog', *Topics in Cognitive Science*, **1**, pp. 292–304.

Gergely, G., and Csibra, G. [2005]: 'The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy', *Interaction Studies*, **6**, pp. 463–81.

Guagnano, D., Rusconi, E., and Umiltà, C.A. [2010]: 'Sharing a task or sharing space? On the effect of the confederate in action coding in a detection task', *Cognition*, **114**, pp. 348–55.

Herschbach, M. [2012]: 'On the role of social interaction in social cognition: a mechanistic alternative to enactivism', *Phenomenology and the Cognitive Sciences*, **11**, pp. 467–86.

Hutto, D.D. [2004]: 'The Limits of Spectatorial Folk Psychology', **19**, pp. 548–73.

————— [2007]: 'The narrative practice hypothesis: origins and applications of folk psychology', *Royal Institute of Philosophy Supplement*, **60**, pp. 43–68.

Ito, T., Tiede, M., and Ostry, D.J. [2009]: 'Somatosensory function in speech perception.', *Proceedings of the National Academy of Sciences of the United States of America*, **106**, pp. 1245–8.

Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Ito, K., Fukuda, H., Kojima, S., and Nakamura, K. [1999]: 'The human amygdala plays an important role in gaze monitoring', *Brain*, **122, pp. 779**–83.

Levinson, S.C. [2016]: 'Turn-taking in Human Communication - Origins and Implications for Language Processing', *Trends in Cognitive Sciences*, **20**, pp. 6–14.

Masangkay, Z.S., McCluskey, K. a, McIntyre, C.W., Sims-Knight, J., Vaughn, B.E., and Flavell, J.H. [1974]: 'The early development of inferences about the visual percepts of others.', *Child development*, **45**, pp. 357–66.

Matheson, H., Moore, C., and Akhtar, N. [2013]: 'The development of social learning in interactive and observational contexts', *Journal of Experimental Child Psychology*, **114**, pp. 161–72.

Michelon, P., and Zacks, J.M. [2006]: 'Two kinds of visual perspective taking.', *Perception & psychophysics*, **68**, pp. 327–37.

Möttönen, R., and Watkins, K.E. [2009]: 'Motor Representations of Articulators Contribute to Categorical Perception of Speech Sounds', *Journal of Neuroscience*, **29,** 9819–25.

Murray, L., and Trevarthen, C. [1985]: 'The infant in mother-infant communication', *Journal of Child Language*, **13**, pp. 15–29.

Nagel, J. [2013]: 'Knowledge as a mental state', *Oxford Studies in Epistemology*, **4**, pp. 273–306.

Pageler, N.M., Menon, V., Merin, N.M., Eliez, S., Brown, W.E., and Reiss, A.L. [2003]: 'Effect of head orientation on gaze processing in fusiform gyrus and superior temporal sulcus', *NeuroImage*, **20**, pp. 318–29.

Pickering, M.J., and Branigan, H.P. [1999]: 'Syntactic priming in language production', *Trends in Cognitive Sciences*, **3**, pp. 136–41.

Pickering, M.J., and Garrod, S. [2004]: 'Toward a mechanistic psychology of dialogue.', *The*

Behavioral and brain sciences, **27**, pp. 169-190-226.

——— [2013]: 'An integrated theory of language production and comprehension.', *The Behavioral and brain sciences*, **36**, pp. 329–47.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y. [2006]: 'Motor cortex maps articulatory features of speech sounds.', *Proceedings of the National Academy of Sciences of the United States of America*, **103**, pp. 7865–70.

Qureshi, A.W., Apperly, I., and Samson, D. [2010]: 'Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: evidence from a dual-task study of adults.', *Cognition*, **117**, pp. 230–6.

Reddy, V. [2008]: 'How infants know minds', Cambridge, MA: Harvard University Press.

Samson, D., Apperly, I., Braithwaite, J.J., Andrews, B.J., and Bodley Scott, S.E. [2010]: 'Seeing it their way: Evidence for rapid and involuntary computation of what other people see.', *Journal of Experimental Psychology: Human Perception and Performance*, **36**, pp. 1255–66.

Sato, W., Yoshikawa, S., Kochiyama, T., and Matsumura, M. [2004]: 'The amygdala processes the emotional significance of facial expressions: an fMRI investigation using the interaction between expression and face direction', *NeuroImage*, **22**, pp. 1006–18.

Schilbach, L., Wohlschlaeger, A.M., Kraemer, N.C., Newen, A., Shah, N.J., Fink, G.R., and Vogeley, K. [2006a]: 'Being with virtual others: Neural correlates of social interaction', *Neuropsychologia*, **44**, pp. 718–30.

Schilbach, L., Wilms, M., Eickhoff, S.B., Romanzetti, S., Tepest, R., Bente, G., Shah, N.J., Fink, G.R., and Vogeley, K. [2010a]: 'Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry', *Journal of Cognitive Neuroscience*, **22**, pp. 2702–15.

Schilbach, L., Eickhoff, S.B., Cieslik, E., Shah, N.J., Fink, G.R., and Vogeley, K. [2011]: 'Eyes on me: an fMRI study of the effects of social gaze on action control', *Social Cognitive and Affective Neuroscience*, **6**, pp. 393–403.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., and Vogeley, K. [2013]: 'Toward a second-person neuroscience', *Behavioral and Brain Sciences*, **36**, pp. 393–414.

Sebanz, N., Knoblich, G., and Prinz, W. [2003]: 'Representing others' actions: just like one's own?', *Cognition*, **88**, pp. 11–21.

Shimpi, P.M., Akhtar, N., and Moore, C. [2013]: 'Toddlers' imitative learning in interactive and observational contexts: The role of age and familiarity of the model', *Journal of Experimental Child Psychology*, **116**, pp. 309–23.

Surtees, A., Butterfill, S., and Apperly, I. [2012]: 'Direct and indirect measures of Level-2 perspective-taking in children and adults.', *The British journal of developmental psychology*, **30**, pp. 75–86.

Surtees, A., Samson, D., and Apperly, I. [2016]: 'Unintentional perspective-taking calculates whether something is seen, but not how it is seen', *Cognition*, **148**, pp. 97–105.

Ten Bosch, L., Oostdijk, N., and De Ruiter, J.P. [2004]: 'Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues', *International Conference on Text, Speech and Dialogue*, Berlin: Springer, pp. 563–70.

Tourville, J.A., Reilly, K.J., and Guenther, F.H. [2008]: 'Neural mechanisms underlying auditory feedback control of speech', *NeuroImage*, **39**, pp. 1429–43.

Tylén, K., Allen, M., Hunter, B.K., and Roepstorff, A. [2012]: 'Interaction vs. observation: distinctive modes of social cognition in human brain and behavior? A combined fMRI and eye-tracking study', *Frontiers in human neuroscience*, **6**.

Umiltà, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., and Rizzolatti, G. [2001]: 'I Know What You Are Doing', *Neuron*, **31**, pp. 155–65.

Van Overwalle, F. [2009]: 'Social cognition and the brain: A meta-analysis', *Human Brain Mapping*, **30**, pp. 829–58.

Vlainic, E., Liepelt, R., Colzato, L.S., Prinz, W., and Hommel, B. [2010]: 'The virtual co-actor: The social Simon effect does not rely on online feedback from the other', *Frontiers in Psychology*, **1**, pp. 1–6.

Watkins, K.E., Strafella, A.P., and Paus, T. [2003]: 'Seeing and hearing speech excites the motor system involved in speech production', *Neuropsychologia*, **41**, pp. 989–94.

Wellman, H.M., and Liu, D. [2004]: 'Scaling of theory-of-mind tasks', *Child Development*, **75**, pp. 523–541.

Wicker, B., Perrett, D.I., Baron-Cohen, S., and Decety, J. [2003]: 'Being the target of another's emotion: a PET study', *Neuropsychologia*, **41**, pp. 139–46.

Yoon, J.M.D., Johnson, M.H., and Csibra, G. [2008]: 'Communication-induced memory biases in preverbal infants', *Proceedings of the National Academy of Sciences*, **105**, pp. 13690–5.