Consciousness and Cognition 36 (2015) 519–531

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Social cognition in simple action coordination: A case for direct perception ☆

Ekaterina Abramova *, Marc Slors

Radboud University Nijmegen, Erasmusplein 1, 6500 HD Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 5 October 2014
Revised 21 April 2015
Accepted 23 April 2015
Available online 21 May 2015

Keywords:
Direct social perception
Action coordination
Affordances
Recursive mindreading

ABSTRACT

In this paper we sketch the outlines of an account of the kind of social cognition involved in simple action coordination that is based on direct social perception (DSP) rather than recursive mindreading. While we recognize the viability of a mindreading-based account such as e.g. Michael Tomasello's, we present an alternative DSP account that (i) explains simple action coordination in a less cognitively demanding manner, (ii) is better able to explain flexibility and strategy-switching in coordination and crucially (iii) allows for formal modeling. This account of action coordination is based on the notion of an agent's field of affordances. Coordination ensues, we argue, when, given a shared intention, the actions of and/or affordances for one agent shape the field of affordances for another agent. This a form of *social* perception since in particular perceiving affordances for another person involves seeing that person as an agent. It is a form of social *perception* since it involves perceiving affordances for another person and registering how another person's actions influence one's own perceived field of affordances.

1. Introduction

Up until a decade ago there was a scholarly near-consensus over the idea that understanding others is an inferential process, called 'mindreading', that can be broken down in two parts. First, understanding the behavior of others is held to involve the attribution of mental states—propositional attitudes, usually—that cause observed behavior. Secondly, mental states are held to be not directly observable. There is a lot of disagreement over the nature of mental states and there are different readings of 'unobservability', to be sure (Bohl & Gangopadhyay, 2014). But accepting a version of both points of departure is accepting that understanding others is inescapably inferential: to understand the actions of others we need to go beyond what is perceivable and make inferences about their hidden mental causes. For a long time, the debate on social cognition concentrated on the nature of these inferences; this is what is at stake in the ongoing discussion between theory-theorists and simulationists. Recently, however, some philosophers have challenged this inferential nature of at least some social cognitive processes. On the one hand, phenomenologists question the unobservability of mental states and insist that we can directly perceive e.g. the other's emotions or basic intentions in facial expressions, voice intonations, gestures and bodily postures (e.g. Gallagher, 2004; Gallagher, 2008; Gallagher & Zahavi, 2008; Zahavi, 2005; Zahavi & Parnas, 2003). On the other hand, the idea that understanding others always involves ascribing full-blown mental states has been

☆ E. Abramova's work on this paper was funded by the Netherland Organization for Scientific Research (NWO), project number PGW-12-32.
* Corresponding author.
E-mail address: m.slors@ftr.ru.nl (M. Slors).

http://dx.doi.org/10.1016/j.concog.2015.04.013
1053-8100/

questioned (e.g. Hutto, 2008; Hutto & Myin, 2012; Kiverstein, 2011; Ratcliffe, 2007). Perceiving contextualized behavior is thought, in some cases, to constitute a non-inferential understanding of a basic kind of intentionality that does not involve propositional attitudes (cf. Hutto, Satne, & Satne, 2015). Thus, in different ways, a case has been made for the notion of 'direct social perception' (DSP).

DSP is usually associated with the idea that understanding others and successfully interacting with them in daily life are inextricably intertwined (Gallagher, 2003). More traditional (cognitivist) approaches would agree only partially. They would agree that understanding others is a necessary *precondition* for successful social interaction but deny that successful interaction necessarily *equals* proper understanding of the other. In this paper we focus on situations in which understanding others serves the purpose of coordinating actions in simple tasks that require two people to collaborate. Such situations are usually described by philosophers in terms of shared intentions. We shall not concentrate on the nature of shared intentions (see Bratman, 1993, 2014; Gilbert, 1989, 2003; Searle, 1995; Tuomela, 1992, 2007). Rather, we shall focus on the kind of understanding of others that guides joint action coordination: who will do what in order to achieve the shared goal? In order to negotiate the roles played in the pursuit of a common goal, one needs to grasp the intentions in the actions of others so as to determine one's own actions, either to complement or to influence the other's actions. This is an extremely common form of social cognition. It is the day-to-day understanding of others involved in jointly tidying up a room, jointly fixing a bike, maneuvering a couch through the house together, or in cooking with a partner. In such activities, settling the 'who does what' question is usually driven by a non-verbal understanding of what the other is up to in conjunction with some grasp of how a single task can be executed jointly. The question we are concerned with in this paper is whether such understanding involves inferential 'mindreading' or whether it can be conceived of as a form of direct social perception. We shall argue that both options are theoretically feasible but make a case specifically for direct perception.

In his defence of DSP, Gallagher (2008) insists that social perception is *smart perception*, i.e. perception that is informed by previous experience, emotions and relevant situational context. Clearly, the combination of such different sources of information in perceiving people is a complex process, but Gallagher insists this is not mindreading. All that is required is perceiving the other person in a situation in which one is also engaged and having a vast store of experience of social interaction. Given these conditions, intentions and emotions of the other are directly perceived. While highly informative, such accounts of social perception have not been sufficiently fleshed out yet when it comes to the perceptual mechanisms involved in perceiving others in their world-context. Nor do they show how social perception figures in interacting with others in action coordination tasks. We will argue that the notion of affordances is helpful in filling this gap.

More specifically, we shall propose that successful coordination of actions in simple cases of joint agency need not be understood in terms of the ascription of mental states that must be inferred from perceived behavior. Instead, it can be understood in terms of perceiving the actions of one's cooperation partner and/or her positioning in a shared environment as directly determining one's own "field of affordances"—the array of action opportunities that one is responsive to. This can be understood as perceiving intentions *in* (rather than 'behind') the actions of others or *in* their being poised for action. But 'perception' is understood in enactivist terms here. Thus, perceiving intentions in the actions and postures of others *means* perceiving them as co-determining the perceiver's own possible action-array so as to collaborate efficiently. This *is* a kind of social cognition in the sense that this involves seeing the other person as an intentional being (rather than a mere physical object). Yet it does not involve the ascription of mental states as hidden causes of action—all intentionality is 'out in the open'. To further clarify our account we will augment it with a preliminary version of a computational model cast within the framework of Dynamic Field Theory (Erlhagen & Schöner, 2002; Spencer, Perone, & Johnson, 2009). We modify the DFT approach so as to (1) incorporate the notion of a field of affordances and (2) make room for embodied cognitive mechanisms spanning multiple agents. This is important since one of the weaknesses of many DSP proposals is that while criticizing an account of social cognition that can be modeled in meta-representational terms, no alternative style of modeling is proposed.

The paper is set up as follows. In Section 1 we briefly present the notion of simple action coordination and the widely accepted idea, well expressed by Michael Tomasello, that such coordination requires recursive mindreading. In order to facilitate our argument for a DSP account of action coordination we will distinguish two types of coordination in Section 2: distributive and contributive action coordination. Roughly, the former is the kind of coordination that involves roles that can be carried out relatively independently of each other, whereas the latter requires e.g. acting on the same object together as in the cases discussed by Tomasello. Distributive action coordination will figure as a stepping stone in our argument. In Section 3 we will claim that the phenomenon of distributive action coordination can be described in terms of one agent's 'field of affordances' being determined by the other agent's actions. In Section 4 we shall describe how DFT modeling can capture this type of coordination process, using an already implemented example from robotics. Up till then we will not have made a case for a DSP alternative for Tomasello's recursive mindreading, which concerns contributive action coordination. In Section 5, however, we will argue that the action-coordination processes described in Section 4 can be scaled up to cover cases of contributive action coordination. Contributive action coordination, we shall argue, can be understood in terms of one agent's field of affordances being determined by the perceived affordances for the other agent. We give reasons to be hopeful that future DFT modeling will be able to implement this type of process.

2. Action coordination and recursive mindreading

The idea that coordinated joint action requires mindreading on the part of the collaborators is widely shared. It is intuitive to attribute a three-stage structure to coordinated joint action. First, collaborators must share an intention; if there are two collaborators, they both intend to do *x* together. As mentioned in the introduction, there are various ways to conceive of this first stage. This is the shared intention stage. We will concentrate on the other two steps, however, that constitute the action-coordination stages. Secondly, there has to be something like a shared action plan. This is crucial for our discussion. Tomasello et al. express this notion as follows:

> (...) both collaborators must choose their own action plan in the activity in light of (and coordinated with) the other's action plan: my role is to hold the box steady while you cut it open. This requires that each participant cognitively represent both roles of the collaboration in a single representational format – holistically, from a "bird's-eye view," as it were – thus enabling role reversal and mutual helping.
>
> [Tomasello et al., 2005, 681]

One complicating factor here is that there may be more than one set of roles that are jointly sufficient to carry out the shared intention. We shall ignore this complication for now. The main point is that, intuitively, it seems that both collaborators need to be aware of a (the same) set of roles that are jointly sufficient to attain a shared goal. Given this shared action plan, a third step required for coordinated joint action would be to assign roles: who is going to do what? This step will be our main concern at first (though we will argue that there is a need to reassess step two later on).

In order to assign roles, we need to take into account the various skills of the collaborators, their positions in the shared physical environment, their access to e.g. tools, etc. But apart from such practical considerations collaborators need to assess, mutually, each other's preferences and inclinations to adopt a certain role. How? Tomasello, writing about "(...) multiple individuals pursuing a joint goal – and all knowing together that they are doing so – with inter-related, complementary roles" answers this question as follows:

> This kind of collaboration requires (...) skills and motivations for shared intentionality, including as the basic social-cognitive pre-requisite, recursive mindreading".
>
> [Tomasello, 2008, p. 173]

The idea here is that collaboration and role distribution requires knowing what others are up to, and knowing that necessarily involves accessing their inclinations, preferences and intentions—i.e. their minds. Though Tomasello is not always clear on this, it is not uncommon within this line of thinking, to suggest (as e.g. Bratman, 1993 does) that collaboration requires mutual knowledge. Thus, collaboration does not only require that I know what you're up to, but also that I know that you know that I know what you're up to, that you know that I know that you know what I'm up to, etc. All this clearly requires both of us to access each other minds. Hence mindreading is deemed to be a "basic social-cognitive prerequisite" (Tomasello, 2008, 173) for collaborative joint agency. This mindreading is recursive because I am assessing your inclinations to act thus or so while you are assessing mine. Tomasello:

> (...) any time we create a joint goal, this must involve a kind of negotiation that inherently involves such mental coordination because I only want to engage in the collaborative activity if you do also (and you feel the same about me), and so we must both assess the other's propensities, which depend on their assessment of our propensities, and so on".
>
> [Tomasello, 2008, 188]

Elsewhere, Tomasello speaks of 'intention reading': "The *key cognitive substrate* required for skillful collaboration is the ability to read intentions." (Tomasello, Carpenter, Call, Behne, & Moll, 2005, 687; italics ours) Intention reading crucially goes beyond appraising situated actions. The latter may be sufficient for competitive behavior, Tomasello argues, but not for collaboration.

Intention reading involves a "means-end analysis of the observed behavior" (Tomasello et al., 2005, 687). The 'ends' cannot be perceived or observed; they rather need to be inferred. Starting from observed situated behavior, inferential cognitive processes produce hypotheses about the intentions and inclinations of others, which in turn imply some understanding of and predictions about the behavior of others required to adjust one's own behavior for the purpose of collaboration. While this may involve the application of 'theory of mind' abilities, according to some, this is not necessarily implied by the suggestion that intention reading is an analytical, inferential affair. Tomasello states that his proposal is a close kin of the 'theory of mind' option (2005, 690), yet according to him the ability to simulate others suffices (see also e.g. Bekkering et al., 2009; Eskenazy, van der Wel, & Sebanz, 2012; Sebanz, Bekkering, & Knoblich, 2006). But here again, such simulation is explicitly meant to go beyond mere bodily simulation as Tomasello and colleagues write in connection with this that "infants begin to understand particular intentional and mental states in others only after they have experienced them first in their own activity and then used their own experience to simulate that of others" (Tomasello et al., 2005, 688). We need to be aware of our own mental states and their relation to our own behavior before we can read the intentions of others through simulation. In

Searle's terminology (Searle, 1983) Tomasello's simulation is not only aimed at acquiring knowledge of the *intentions-in-action*[1] of other people, but also of their *prior intentions*—intentions that are hidden rather than in plain view. Tomasello's proposal here is completely in line with simulationists such as Goldman (2006). On Goldman's view, ascribing intentions (and other mental states) involves an 'inference from me to you' (as Gordon puts it, Gordon, 1995). Simulation in this sense is just as much inferential as the use of 'theory of mind' abilities.

To be sure, the inferences at play here are not suggested to consist of conscious reasoning processes. Just like our wielding a 'theory of mind', the application of simulation routines aimed at understanding others is usually depicted as a sub-personal cognitive process that should not be confused with the personal-level phenomenology of interacting with others (Herschbach, 2008; Spaulding, 2010). The claim that the kind of understanding of others that is involved in collaborative action coordination is inferential is a claim about (sub-personal) cognitive processing only. Likewise, the concomitant idea that the intentions of others are said to be 'hidden' behind observed behavior on views such as Tomasello's and Goldman's needs to be taken only as a claim about the inferential rather than perceptual nature of the social-cognitive processes involved, not as a claim about the phenomenology of collaboration or the nature of the ascribed states (Bohl & Gangopadhyay, 2014).

3. Intermezzo: Two kinds of action coordination

Not everyone is convinced that the social cognitive processes underlying simple action coordination are inferential. Some conceive of the simulation involved in terms of 'common coding' of perception and action without implying an inference from me to you (Bekkering et al., 2009; Eskenazy et al., 2012; Sebanz et al., 2006). Others stress that the high-level mindreading involved in sharing intentions should be complemented by an 'alignment system' consisting of all kinds of low-level embodied cognitive processes that facilitate interaction and collaboration (Tollefsen & Dale, 2012). We wish to take this line of thinking one step further and investigate whether it is possible to conceive of the social cognitive processes involved in simple action coordination in perceptual rather than inferential terms. In order to present our case as clearly as possible it is useful to make a distinction between two kinds of simple action coordination.

Much of the literature on simple coordinated joint action involves cases in which dyads are required to act simultaneously on the same physical object. In Tomasello's example discussed above, two people act on a box; one holds it steady while the other cuts it open. Other such examples include carrying pieces of furniture together (McCabe, Houser, Ryan, Smith, & Trouard, 2001) or jointly lifting planks while each participant touches only one side of the plank (Isenhower, Marsh, Carello, Baron, & Richardson, 2005). In such cases coordinated joint action is usually required to start simultaneously. The actions of the individuals involved contribute to one kind of handling of a physical object. Thus, participants need to track the actions of the other participant in order to execute her own part. We can refer to this kind of action coordination as contributive action coordination.

There are also cases of simple coordinated joint action in which participants are not required to act on the same object. Think, for instance, of jointly cleaning up a room. Here one person may start vacuum cleaning. The next person chooses a different chore and so on until all roles required to reach the joint goal of cleaning up the classroom are taken. Or think of two waste collectors stepping off a garbage truck and walking toward a stack of bin bags. As one walks toward the left-hand side of the stack, the other will facilitate coordinated action by approaching the right-hand side. In such cases it is usually not necessary to start acting simultaneously. This is because the overall joint task can be broken down into a set of actions that can be executed by individuals without requiring the help of others and often without constantly keeping track of the other's actions. We shall refer to this kind of action coordination as distributive action coordination.

A prototypical instance of simple distributive action coordination is the following. Two people, one of whom is an accomplice to an experimental set-up, are given the task of sorting a bunch of marbles by color. There are 15 blue and 15 white marbles, mixed, in one bowl. The accomplice will start sorting by taking an empty bowl from a stack of bowls and consecutively taking 3 blue marbles and putting them in the empty bowl. She will then ask the other participant to help her sort the marbles. In an informal pilot study with 40 trials of this scenario we found that all of the non-accomplice participants would start by putting white marbles in another empty bowl. They opted for the most straightforward complementary action instinctively. During the trials, some participants changed strategy. Sometimes participants took blue marbles out, which was always due to a temporary inability to easily access white marbles in the mixed-color bowl. Sometimes participants started taking a bunch of mixed-color marbles in one hand, using the other hand to distribute them to the single-colored bowls.[2] This is a case of distributive action coordination; actors act on different physical objects and one actor starts with a single course of action to which the other actor can adapt (in different ways).

[1] On Searle's account intentions in action have (propositional) content. Later on we shall compare intentions in action with Dan Hutto's notion of 'intentional attitudes', which are meant to be content-less. The propositional content of intentions in action do not play a role in our argumentation.

[2] This also boils down to complementing the actions of the accomplice, since now each of the participants works with their own bunch (in our case the accomplice would keep picking blue marbles only). In order to determine whether other divisions of labor may emerge, how many participants opt for which division and under what circumstances they change strategy, this experiment ought to be conducted under controlled circumstances. For our purpose, however, such exact data are not required. All we need is the scenario and the fact that people are inclined to complement each others actions in different ways. Note that participants could also have joined the accomplice in picking out the blue marbles. This would have resulted in achieving the set goal—sorting marbles by colors—even quicker than by complementing each others actions.

The distinction between contributive and distributive action coordination is not meant to be black-and-white. In fact, Tomasello's jointly opening a box example may be considered an in-between case. For on the one hand it involves jointly acting on the same object which requires keeping track of each others actions at least in the first stages of the operation. On the other hand, there is a sequence to the roles (first one has to pick up the box, then the other can start cutting) which means that the participants need not start simultaneously and which also limits the extent to which they need to keep track of each others actions. The distinction between contributive and distributive action coordination can best be conceived of as a gradual one. It will be useful for our discussion, though. For the case for a non-inferential account of the kind of social cognition involved in unambiguous distributive action coordination can be made relatively easily. This is what we shall start out with. The model we shall develop for distributive action coordination can then be amended, making it somewhat more complex, so as to be applicable to contributive action coordination.

4. Distributive action coordination: Redescribing the explanandum

One can describe the cognitive processes behind the simple distributive action coordination in terms of 'intention reading' and inference. Take the marble example. On an elaborate reading, the non-accomplice cooperation partner ascribes an intended overall action plan in Tomasello's sense to the accomplice in which different roles are distinguished (one picks the blue marbles, the other picks the white ones) and distributed (she takes care of the blue ones, so I'll do the whites). But it is also possible to let the action plan emerge on the fly, so to speak. In that case, the reasoning pattern may be something like this: "it is reasonable to assume that the accomplice has the intention to go on picking out blue marbles", "the chance of actors interfering with each other's activity will be minimized when I start picking white ones (or when I take my own bunch in one hand instead of picking from the mixed bowl)", so "I'll start picking white marbles (or take my own mixed bunch)."

Arguably, however, ascribing intentions as hidden causes of behavior is superfluous in this case. For instance, the assumption that the accomplice will continue picking blue marbles may be described as a mere inductive inference rather than an inference based on intention reading (see Andrews 2012 for elaborate argumentation to this effect). But even then, the fact that a process can be *described* as an inductive (or, for that matter, as an intention-reading) inference at the personal level, does not imply that the sub-personal cognitive processes that underly it do in fact instantiate something like an inference. Compare: the distribution of sunflower seeds, at the 'macro' level, can be described in terms of the Fibonacci sequence. But this does not mean that this sequence or some mathematical process that produces this sequence is in fact instantiated in the micro-biological processes that determine the distribution of sunflower seeds.

So the idea here is simply that action coordination that can be described as being 'driven' by intention reading-based inferences may in fact be driven by sub-personal processes that do not resemble intention ascription and intentional inference. The personal-level ways in which processes are described are not fixed by the sub-personal cognitive processing (see Slors, 2012). This does not mean that a process that can be described at the personal level in a specific way can in fact be underpinned by *any* type of sub-personal process (compare a computer program such as 'Word' that runs on a Mac and a PC; this fact does not imply that any computer can run Word). The fact that two or more different kinds of sub-personal process can produce similar or even identical behavior at the personal level means only that observation of that behavior does not tell us, yet, which of a limited set of possible sub-personal processes is at play. By *describing* processes in one way or another, at the personal level, however, we do influence our conception of the sub-personal processing at play.[3] In order to argue for a DSP alternative to ascribing intentions as hidden causes of action as a characterization of the sub-personal processing, then, we will argue for a different personal-level description of the explanandum. We should emphasize here that we are merely offering a *description* of the explanandum. We will turn to explanatory models in the next section.

For our redescription we shall make use of the notion of 'affordances'. The term 'affordance' was coined by Gibson ([1979] 1986, p. 127):

> The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment.

Briefly put, affordances are action possibilities offered by the environment to a particular organism. These possibilities are determined both by (physical, social, cultural) features of the environment and by the bodily and psychological features of the organism (Chemero, 2003) as well as features that are determined by the 'form of life' or culture an organism is immersed in. The bodily features here do not just involve body size and shape, but also skills and abilities. Thus, young adults perceive different stair-climbing affordances than equal sized seniors (Cesari, Formenti, & Olivato, 2003). Note that we will use the notion of 'affordances' in a descriptive way only; we make no claims about their ontological status or causal explanatory role.

[3] This point is made at length by Daniel Dennett in his "Three Kinds of Intentional Psychology" (1987, 43–68) where he describes folk-psychology as the source of theorizing about the sub-personal mechanisms underlying our behavior. We concur with Dennett's description of the relation between personal-level folk-psychology and sub-personal psychology but we do wish to emphasize that there are more personal-level ways of describing our cognitive functioning than classic, belief-desire-based folk-psychological ways.

Any environment of any organism can be said to contain a plethora of different affordances. But only a very tiny fragment of those can actually be said to be involved in the actions an organism engages in. It is thus useful to distinguish between what Rietveld and Kiverstein (2014) call the 'landscape of affordances' and the 'field of affordances'. Whereas the former refers to all action possibilities available to a form of life, the latter refers to action possibilities that an organism—a person, say—is potentially responsive to in a given situation, depending on e.g. psychological factors such as (shared) goals or needs. Thus, the field is a situation-specific, individual subset of the landscape of affordances. An important part of the notion of the field of affordances is that it can dynamically change, corresponding to changes in the situation or internal state:

> If a rabbit eats the only carrot available in a certain place, it changes the layout of the (locally present) landscape of affordances. However, as the landscape of affordances changes and the individual's interest in eating diminishes, new possibilities for action show up. Once the carrot has been eaten, the rabbit hole might solicit sleeping, or a place a bit further away might solicit exploring.
>
> [Bruineberg & Rietveld, 2014, p. 3]

That is, the content of the field can change, i.e. which objects are present in one's environment and their arrangement, but so can its soliciting ''profile'' (the saliency of the objects present). The primary reason for talking about a *field* rather than particular object affordances, as is frequently done in cognitive neuroscience or psychology (when e.g. investigating an activation of brain areas related to actions that can be performed with a given object induced by perceiving that object), is that in order to behave adaptively, an organism has to be able to switch between action possibilities available in a given situation, allowing different elements of that situation to solicit actions[4]. This theoretical perspective is consistent with a neurophysiological account of Cisek (2007) that views action as affordance competition, i.e. a continuous interplay between action selection and action specification in any given situation.

The idea we would like to put forward, as an alternative to an inferential description of the cognitive processes underlying distributive action coordination, is that such coordination can also be described in terms of the actions of our cooperation partner co-determining both the selection of our field of affordances from the landscape of affordances and the changes within the field of affordances as coordination proceeds. Briefly put: perception of the other's actions mediates the perceived array of possible relevant actions directed at our environment. What I see my partner doing directly and non-inferentially determines the action possibilities that are salient to me. At a personal-level of description this can be interpreted—at least in very many instances—as perceiving the intentions-in-action (Searle, 1983) of others, as long as we conceive of perception in enactivist terms. Let us break this proposal down in its two components.

First, intentions-in-action are here intended to be of a kind with what Dan Hutto (2008) calls 'intentional attitudes' (see footnote [1]). They consist of the perceivable goal-directedness of people's movements or bodily postures. In line with other DSP proposals (Gallagher, 2004; Gallagher, 2008; Gallagher & Zahavi, 2008; Zahavi, 2005), the idea here is that we actually see intentions without having to infer them. It is important to stress that the notion of perceiving intentional attitudes is pitched at the personal level of description. Thus we do not reject various proposals for understanding the sub-personal mechanisms underlying the direct pick up of other people's basic intentions. In particular we are not opposed to the many proposals according to which mirror neurons play a major role in such pick up (e.g. di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996; Gallese, 2007). We do, however, think that it is misleading to describe such sub-personal processes in terms that are borrowed from higher-level inferentially characterized cognitive processing. In particular we think it is misleading to think of mirror neuron activity as a form of mental simulation. Rather than as simulation, mirror neuron activity can more neutrally be described as 'neural resonance' (Gallagher, 2007; see however also Slors, 2010) so that the sub-personal processes underlying our picking up on another person's intentional attitudes are not depicted as hypothesizing about another person's inner realm.

Secondly and importantly, perception enactively conceived does not mean adequate representation of the perceived information. Rather, it means usage of the perceived information in driving adequate actions of the perceiver. Thus, perception of another person's intentions-in-action, enactively understood, means picking up on the intentional directedness of the other on a shared environment so as to co-determine the way in which this environment solicits actions by the perceiver. This is, in effect, the idea that the actions and postures of other people can co-determine the selection of our field of affordances from the complete landscape of affordances. The fact that this alternative conception of action coordination can also be described as enactive perception of intentions in action is important. For it underscores the idea that, even though no ascription of hidden mental states is involved, this is still a form of direct *social* perception. Others are relevant in determining one's own field of affordances and hence one's actions as *intentional* agents, not merely as physical objects.

The cases of cleaning a room together or collecting bin bags can easily be described in terms of others co-determining the selection of our field of affordances from the landscape of affordances and the changes within the field of affordances as coordination proceeds. But we shall concentrate on the marble sorting scenario since this is a little bit more complex. The idea is that the initial actions of the accomplice (consecutively taking out 3 blue marbles and putting them in a separate bowl) mediates the array of action possibilities perceived by the cooperation partner. Initially this leads, apparently, to the affordances of the white marbles (being picked out of the mixed bowl and put in a separate empty one) becoming salient for the

[4] Although the field could be said to be composed of individual object affordances, the notion of the *field* helps emphasize that affordances not acted upon at any given moment are capable of retaining residual activation and influencing behavior as the situation unfolds.

cooperation partner. Before discussing the possibility of a change of strategy, it is crucial to stress that on our redescription all we need to fix the field of affordances for the cooperation partner (apart from the physical features of the scenario) is the actions of the accomplice. There is no need to postulate unobservable intentions behind these actions. Indeed, as argued above, it isn't even necessary to assume inductive inferences. All we need is the expectation that the accomplice will go on doing what she is doing.

Here it may be argued that such an expectation boils down to a prediction and that such a prediction would require some form of mindreading or the ascription of intentions as hidden causes of action. We disagree. At least at a personal level of description it is alright to translate an expectation into a prediction. But to understand the prediction as being based on or grounded in the ascription of a hypothetical (inferred but not observed) intention is entirely superfluous. For what would be the reason to postulate that intention, *other than the expectation that the accomplice will continue doing what she does*. As Ryle (1949) showed in great detail, we are often so captivated by a causal (para-mechanical) intentional picture of human behavior that we tend to lose sight of the fact that the ascription of most mental states is not just based on observed behavior, but that it is often a characterization of observed behavior. To expect others to continue doing what they do in the absence of a reason to change one's behavior may quite simply be a psychological fact (as is e.g. postulated by the current 'predictive coding' paradigm (Hohwy, 2013)); we may well form this expectation, even if there is no underlying inferential justification. For psychology should not be mistaken for epistemology.

Equally important is the fact that on our affordance-based redescription of the explanandum, there is no need for a discrete consideration about being in each other's way. Seeing the accomplice pick blue marbles and expecting her to continue doing so will simply eliminate the blue marbles from the collaborator's field of affordances on the description we propose. To say that this elimination is grounded in an inference based on a preference not to be in each others way is to confuse the logical justification of an action for the sub-personal cognitive mechanisms at play.

The description of the explanandum we would like to put forward suggests a simpler and less demanding cognitive process (if it is that at all) than Tomasello's shared action plans. This is perhaps best illustrated by considering the questions why and how people change strategy in the marble sorting scenario. The most common change we observed was the change from a "complementary color strategy" (one in which one person picks out e.g. white marbles while the collaborator picks out blue ones) to a "take your own batch strategy" (taking a batch of mixed marbled in one's own hand and sorting them into bowls before taking a new hand). Typically this change is initiated when turn-taking is not going smoothly and the participants hands are in each others way in the bowl. When consistently taking a batch of mixed marbles from the bowl by one participant and taking away one color only by the accomplice results in an overrepresentation of one color (blue if the accomplice picks white) this prompts some participants to return to the one-color-only, one-marble-at-a-time strategy. These are just examples of changes of strategy and what prompts them. What we propose is that these changes of strategy are direct responses, unmediated by inference, to changes in the perceived field of affordances. When turn-taking is not going well and hands are in each others way, the bowl with marbles will no longer be perceived as easily affording taking out one blue marble at a time; instead it will be perceived as more easily affording taking out a batch of mixed colored marbles. The change of strategy is an immediate response to this change in the participant's perceived field of affordances.

Of course we can also describe the change of strategy in terms of a change of overall action plan, based on inference and ascription of intentions. The participants hands being in each others way may be conceived of as providing good *reasons* to opt for an alternative overall action plan that may be deemed more efficient or easier to carry out. We have no knock-down argument against this option. But we do wish to point out that this option involves the cognitive processing or representing of two elements that are not involved in the affordance-based reconstruction of what is going on in this type of action coordination.

Firstly, and most conspicuously, an overall action plan involves the actions of the other agent—the accomplice—as well as the actions of the participant. Thus, if a change of strategy involves a change of overall action plan, this means that the anticipated actions of the accomplice are also represented in the new overall action plan. In the marble example, the actions of the accomplice remain the same throughout the whole experiment. That is, when a participant changes strategy and switches from picking blue marbles, one at a time, to taking a whole handful of mixed-colored marbles, the accomplice will continue to pick blue marbles, one at a time. The accomplice's continued actions will fit an overall action plan in which the participant takes a handful of mixed marbles just as well as an overall action plan in which the participant picks one color only, one marble at a time.

Is it necessary for the participant to represent the anticipated actions of the other? On an action-plan account of coordination it is. It is at least necessary to take into account the anticipated future actions of the accomplice as one part of an overall action plan that promotes the shared goal and does so while not interfering with the participant's intended future actions (in cases of distributed action coordination, non-inference is the most crucial constraint on a collaborator's anticipated actions; in the cases of contributive action coordination, the constraints are much more elaborate—see below). On an affordance-based account of action coordination explicitly representing anticipated future actions of one's collaboration partner is not necessary. For on this account, action coordination is driven by 'on-line', direct adaptation to changes in the agent's environment. If the accomplice changes her actions, this will be a prompt to re-attune the actions of the participant. But when there is no change, the actions of the accomplice are 'merely' a part of the participant's environment, to which she is directly responsive, not a part of the participant's mental representation thereof.

Secondly, an action-plan based account of changing strategies involves the weighing of options. One action plan is abandoned in favor of another. This change is not just prompted, there is a reason for it. Obviously, this reason need not be considered consciously; it is clear that participants do not need to deliberate explicitly about their strategies. But on an action-plan based model of changing strategies, one plan must be deemed better, at some level of cognitive processing, than another. And for this, *both* must be represented and compared, at least during some brief period prior to the time of switching. Again, such double representing and comparing is not required by an account of strategy changes that is based on online adaptation, such as the affordance-based account we put forward. On this account, changes in the environment may simply tip some perception–action balance in the participant.

So, the example of strategy changes demonstrates at least two ways in which a Tomasello-like action-plan account of action coordination is cognitively more demanding than the affordance-based account we proposed. This, we submit, is a good reason to see whether an affordance based, direct perception account of action coordination can be developed further.

5. Modeling distributive action coordination on the affordance-based description

Distributive action coordination can be modeled in meta-representational terms when it is described in terms of intention ascription, shared action plans and inferences. What we would like to show is that the affordance-based redescription of the coordination process—the field of affordances of one collaboration partner being determined by the actions of the other collaborator—can be modeled equally well in ways that can e.g. allow robots to collaborate. We start by describing the modeling framework that we adopt for our purposes, namely the Dynamic Field Theory (DFT). We choose this framework for its compatibility with our basic claims as well as its high metaphorical and visual value in elucidating the view on action coordination that we propose.

DFT is a particular instantiation of a dynamical systems approach to cognition in which emphasis is put on (1) the temporal unfolding of cognitive processes and (2) the emergence of stable behavior from a multiplicity of lower-level component processes each governed by their own dynamics. The basic currency of the DFT framework are dynamic neural fields (DNF). These are formalizations of the distribution of activation in populations of neurons relative to continuous metric dimensions in the organism's environment. The dimensions can be anything that is of interest to the modeler but typically the focus is on spatial locations and movement directions. Each site in a neural field corresponds to a particular value of the dimension, for example a direction in which an agent can move their hand (see Fig. 1). Activation is "the certainty with which some bit of information is currently known to the nervous system" (Spencer et al., 2009, p. 91). In the case of a movement field, a high activation value of a particular site in the field would mean that a movement in the direction corresponding to the site is likely.

The fields are dynamic because activation levels change with time according to intrinsic dynamics as well as inputs from other parts of the system (e.g. senses) which are themselves represented as fields. The resulting patterns of change are partly governed by interconnections between different neurons and lead to behaviorally relevant phenomena, for example an ability to self-sustain activation in the absence of external input (which corresponds to working memory).

A paradigmatic example of the use of the DFT modeling approach is an account of infant perseverative reaching, advanced by (Thelen, Schöner, Scheier, & Smith, 2001). In the task, an infant is asked to retrieve an object from one of two containers after she observed the experimenter hide it in one of them. In the first couple of trials, the object is consistently hidden in container A and the infant (from around 7 months of age) is able to retrieve it. Curiously, if at this point the toy is hidden in the second container, B, the infant (up to around 12 months) will still reach for the first container, exhibiting the so-called A-not-B error.

A DFT formalization was proposed to account for the error itself, its dependence on a variety of factors, and the developmental change that allows older infants to perform correctly. It was meant as an alternative to the standard explanation that the error occurs due to some lack of (or lack of access to) relevant knowledge structures about hidden objects. Instead, the DFT explanation focuses on how the infant's behavior emerges from a coupling between different neural, bodily and environmental resources. All of these resources are represented in the DFT model as fields: (1) input fields that represent the task environment with two containers and the actions of the experimenter, (2) the memory field that contains a trace of previous actions (3) the output field that represents where the infant will reach (the movement planning field). The age effect is explained by a model parameter that expresses cooperativity between field sites and the resulting ability to form self-sustaining peaks of activation. Put simply, younger infants have lower cooperativity of the movement planning field and therefore their memory of the experimenter's cue to the B-container decays during a time between the cue and response and the movement is primarily driven by strong memory of previous reaches.

Since the early focus on modeling simple behaviors such as saccadic movements (Kopecz & Schöner, 1995) and reaching (Bastian, Riehle, Erlhagen, & Schöner, 1998), the DFT framework has expanded to incorporate models of working memory (Johnson, Spencer, Luck, & Schöner, 2009), sequence generation (Sandamirskaya & Schoner, 2008b), and even preliminary attempts at capturing turn-taking in communication and spatial language (Sandamirskaya & Schoner, 2008a). What is of primary interest to us here is the use of dynamic fields in simulating action coordination between two agents. It is easy to see that in order to capture such interaction, what needs to be added to individual fields of each of the agents involved, is a coupling between their fields. One might ask, however, if such a coupling amounts to shaping each other's field of affordances as our theory would expect. The answer, obviously, depends on whether we can find affordances in the dynamic fields employed by DFT.

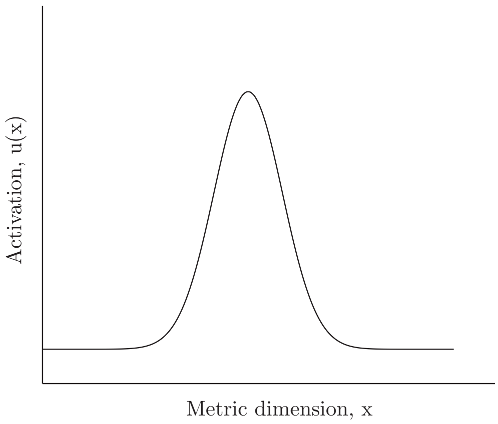

Fig. 1. Dynamic neural field example.

It is noteworthy that already in the commentary to the BBS target article on the DFT account of the A-not-B error (Thelen et al., 2001) a suggestion appeared that the movement-planning field postulated by the authors can be viewed as an affordance field, in which the dimension would be based not on mere spatial metric dimension but rather on types of actions that can be performed on objects (Glenberg, Cowart, & Kaschak, 2001). The details of the commentators' argument are not relevant to the present discussion but we would like to note several things. First, Thelen and colleagues define their task input field as "the features of the world which constitute the behavioral alternatives within the intentions of the actor" (p.18), which is directly compatible with Rietveld's notion of the field of affordances. Second, the authors in response to the commentary also note that while in the A-not-B experiment the task field is assumed to be neutral and novel to all infants, in reality it is of course landscaped through experience, which then acts as a pre-shape for the movement field - a feature that would correspond to the relational structure of affordances (their dependence on the agent's action possibilities). Third, due to the interaction of memory and task input, the movement planning field can contain a number of sub-threshold solutions which can become active if the situation is changed – a feature corresponding to the flexible switching between action possibilities within a field. And finally, in simple tasks such as sorting marbles, there is no need to go beyond a spatial notion of affordances because locations of particular marbles and the target bowls are the only relevant dimensions of the task input. However, more complex affordances could in principle be represented with multi-dimensional fields if the performance of the task would so require (e.g. if one agent had to perform a power grip while another a precision grip). To sum up, there is no need to represent a separate "dynamic neural field of affordances" in order to capture the philosophical notion of the "field of affordances". Rather, the properties that are definitional of the latter can be found in different properties of different dynamic neural fields relevant to action and the way they interact with each other.

We suggest that it is possible to model distributive action coordination in cases like the marble task in terms of dynamic fields. We are currently constructing an experimental paradigm that will enable us to test our predictions in a more controlled fashion. However, a similar model for distributive joint action already exists. (Bicho, Hipólito, Louro, Cambon, & Erlhagen, 2006) use a DFT model in two robots that jointly search for objects in a cluttered environment and transport them to one site in the most efficient way possible. The robots do not communicate explicitly but joint action is accomplished by a combination of memory and prediction abilities of the robots. Each of the robots possesses the following dynamic fields: (1) an STS-like field—i.e. a field that represents the kind of information normally processed by the superior temporal sulcus in the brain—that represents the robot's perception of the motion of his partner, (2) a working memory field that keeps a memory trace of objects in the environment, (3) a goal field that implements the guidance of upcoming action sequence and (4) a "premotor cortex" field that guides the action.

In an illustrative example, one robot, R1, can perceive two target objects, T1 and T2, chooses the closer one (T2) and starts moving toward it. Because of its working memory, R1 is able to remember the location of T1 even when it gets occluded by an obstacle (recall that the robots are moving in a cluttered environment). The STS field of R1 takes in the information about the movement of the second robot, R2 and "predicts" that it is moving toward T2. Crucially, this prediction does not involve a 'means-end' analysis (as in Tomasello's intention-reading). This prediction inhibits R1's decision to move to T2 and it instead switches to moving toward T1 (based on the remaining memory trace) and both robots end up accomplishing the task efficiently. This is distributive action coordination (see the source for detailed figures of the model).

The dynamics of all the fields above and the way they lead to behavior can be *described* in terms of inferences, i.e. R1 inferring that R2 intends to go for T2, combining it with some encoded information that the task is to collect objects efficiently and therefore deciding to go for T1. We would like to argue here, however, that a description in terms of fields of affordances being affected by the other's actions is more accurate and much more in line with the actual features of the model. In particular, the model contains no discrete elements that would constitute a means-end analysis either with respect to the other robot's action or with respect to its own actions. By contrast, R1's action field at the beginning of the discussed scenario can very naturally be described in terms of two affordances - moving toward T1 or T2. Due to the inhibitory connection between the perceived target of the R2's action, the tendency to move to T2 decreases and instead T1 becomes more salient. In other words—as we proposed in the previous section—R1's field of affordances is mediated by R2's actions.[5]

6. Contributive action coordination

So far it can be argued that we have not really been discussing the domain of coordinated action that requires real social cognition. That is, if it is possible to accomplish an overall task without constantly tracking the other agent, then maybe simple social perception is enough. It may be argued, though, that once we really need to act together with the other, that is when inferences make a comeback. This is where we turn to *contributive* action coordination (see Section 2), in which agents need to act on the same object. We will first extend our affordance-based view on distributive action coordination to such cases and describe what kind of social perception is involved. Then we shall make some suggestions for future research purporting to demonstrate that this kind of social cognition can be also captured by a DFT model.

In Sections 3 and 4 we argued that distributive action coordination may be subserved by the fact that one agent's field of affordances is altered by the actions of another agent. That is, the actions of another make some affordances in my own field vanish or decrease in salience. In some cases of contributive action coordination a similar mechanism may be at play. For instance, if one person already starts lifting one end of a table she thereby makes the other end afford table lifting in a joint action for another person. But in many cases of contributive action coordination people are required to start acting simultaneously. In a paradigmatic experiment by Isenhower et al. (2005), for instance, people were requested to lift planks appearing on a conveyer belt. They were only allowed to touch the end grain sides of the planks. Some planks were small enough to be lifted by one person, others could only be lifted when two persons would each touch a different side of the plank. Lifting a plank together in this way is certainly a case of coordinated action. But the coordination cannot be due to one person's *actions* influencing another's field of affordances. For the lifting has to commence simultaneously. It can also not be described in terms of *limiting* the actor's affordances as the presence of another person provides me with a possibility of lifting the plank that I did not have on my own. How, then, can such cases of contributive action coordination be fitted into the affordance based scheme we would like to propose?

What we would like to suggest is that in such cases of contributive action coordination it is not the *actions* of others that affect another person's field of affordances, but the perceived *affordances for the other* (crudely put; see below for a more nuanced picture). Furthermore, the influence of perceiving affordances for another person is not *limiting* but *augmenting* in nature. Thus, in the plank-lifting case, the field of affordances of a person standing at one end of a plank is changed when she sees another person at the other end of the plank. The plank becomes 'jointly liftable' for her because she recognizes the availability of exactly the same affordance for the other person as well. In such a case the slightest motion of one of the people involved might trigger the joint action of lifting the plank (depending of course on whether there is a shared intention to do so—remember that we are only talking about the action coordination phase). Or consider Tomasello's box-opening scenario. In that scenario it is apparently the case that (for some reason not disclosed to us by Tomasello) the box can only be cut open when someone holds it. Thus, even though Thomas has a knife, the box by itself does not afford opening to him. But now imagine that Thomas sees Ellen standing behind the box. Thomas' ability to recognize that the box affords lifting to Ellen changes his field of affordances: now the box is perceived by Thomas as affording being cut open in a joint action.

This proposal builds on the idea that we can perceive affordances for others, but in a crucial sense it goes beyond that idea as well.[6] Let us first discuss the possibility to perceive affordances for others, a possibility that was already recognized by Gibson at the inception of the concept of affordances:

> The child begins, no doubt, by perceiving the affordances of things for her, (...). But she must learn to perceive the affordances of things for other observers as well as herself.
>
> [Gibson, 1979/1986, p. 141]

Since this suggestion, there has been a growing body of literature showing that people are able to judge affordances for others almost as accurately as affordances for oneself. The concern in this strand of ecological psychology has been with investigating what information is available in movements, bodies or positions of others in the environment such that it can provide a perceptual basis for social judgments (McArthur & Baron, 1983; Runeson & Frykholm, 1983). For example, people are able to judge whether observed people are able to sit on a chair based on perceived relationship between the actor's leg length and the seat height (Stoffregen, Gorday, Sheng, & Flynn, 1999). From a perspective of ecological psychology, information about action possibilities for others is directly perceived by the pick up of invariants in the perceptual array. More recent accounts invoke an 'affordance-mirroring mechanism' to account for the same ability (Costantini & Sinigaglia, 2011). Which of the accounts is closer to reality is an open empirical issue but we would argue that both are compatible with our view as long as they are conceived of as primarily perceptual processes.

---

[5] In the marble task, the same effect would correspond to moving to a different-colored marble in any single move. We would predict, however, that consistent behavior of the partner - R2 always moving to a particular target or a participant in the marble task always choosing the blue marbles - would lead to a persistent bias in the movement perception field of R1, which would in turn produce a persistent bias in its own movement planning field to respond in a complementary manner.

[6] We thank an anonymous referee for pointing this out.

The notion of perceiving affordances for others can be used to extend our distributive action coordination account: one's field of affordances can be shaped not merely by perceiving actions of others but also by perceiving affordances for others. For example, perceiving that one of the parents in the classroom-cleaning example is standing right next to the vacuum cleaner will bias my decision to wipe the desks instead. This would be the case of (distributive) action coordination in which affordances of the other (not their actions) limit my own action possibilities. Now, let us conceive of the next step in our spectrum of action coordination complexity. Imagine the simple case in which affordances for the other person augment the array of actions available to me without my needing to worry about the other's willingness to cooperate, i.e. without the need to assess their intentions. Perhaps an analogy with new arguments that have been made for viewing communication as a type of tool-use may help to convey this idea (Borghi, Scorolli, Caligiore, Baldassarre, & Tummolini, 2013).

Recent evidence (Costantini, Ambrosini, Scorolli, & Borghi, 2011) suggests that sensitivity to affordances of the environment is modulated by whether the objects whose affordances are in question are in one's peripersonal space, i.e. whether they are within the space that can be acted upon (this is more than spatial proximity as an object can be spatially close but behind a glass shield which will not count as peripersonal space). Borghi et al. (2013) review related evidence (e.g. Iriki, Tanaka, & Iwamura, 1996) that the experience of using tools induces an extension of peripersonal space. That is, a distant object can become reachable provided that I have a rake to pull it closer. They then suggest that words can similarly act as tools, i.e. an object can also become reachable if I can ask another person to give it to me. The suggestion is backed up by empirical evidence (Scorolli, Daprati, Nico, & Borghi, 2011) that peripersonal space is indeed extended in the experiment in which people retrieve objects with the help of a tool, a button and an object word label uttered to the observer and the effect is not affected by the type of a tool used. In a similarly fashion, we suggest that we could perceive the plank as liftable-with-the-other. In such a case, the other and her perceivable action capabilities is part of the environment that offers action possibilities to me. We call this case "simple" because the cognitive process we envision here does not even involve perceiving the other as an agent with mental states – he is simply a physical tool that can augment my interaction with the environment. My perceiving him as a useful tool depends on the long history of observing people's bodies interacting with the environment and my own history of using physical tools and people to extend my action possibilities.

Now, clearly, for a more complex case of contributive action coordination something else is required. In the plank lifting experiment, it is not exactly right to say that one plank-lifter's field of affordances is augmented by the affordances for the person standing at the other end of the plank in the simple tool-use manner. For the affordance to lift a plank for *that* person also depends on the cooperation of the plank-lifter. Now we might think that there is a recursive process in which one person sees the affordance for another to touch one end of the plank. This augments her own field of affordances. For now her touching the other end of the plank would allow her to start lifting it. She can then acknowledge that this, in turn, will augment the affordances for the other person as the plank now affords lifting to her as well. Thus, we can imagine a recursive process of perceiving affordances-for-others that bootstraps a dyad into contributive action coordination, in which the actions of the two agents are mutually dependent.

There are two considerations that count against this option. For one thing, this way of conceiving affordance-based action coordination closely resembles the recursive mindreading approach advocated by Tomasello. Thus it becomes less than obvious that this kind of affordance-based contributive action coordination is a real alternative to Tomasello's proposal. For another, the recursion process is entirely conditional. The perceived affordances for one person hinge not just on this person perceiving affordances for someone else, but also in assuming that this other person will act on these affordances. I can see that *if* someone else touches a plank on one side, my touching the other side allows both of us to lift the plank. But if the other person does not act, there will not be a change in my field of affordances. Thus the bootstrapping process will only be initiated under the assumption that the other person will act on the simple affordance to touch a side of a plank. But why should she do this if the benefit of that action will become apparent only after the bootstrapping process is completed?

Perhaps it is possible to construe this seemingly recursive process in a way that avoids the problems above. For example, by building in the assumption of the other's cooperation into the shared intention of performing the action together. However, we also wish to consider another, more radical alternative by recognizing the possibility of perceiving affordances for dyads (or, for that matter, for groups).

Consider that in their interpretation of the plank lifting experiment, Sebanz and colleagues note that the finding that people switch from lifting planks alone to joint lifting "as a function of a *pair's* mean arm span" (Sebanz et al., 2006, p. 73, emphasis ours) suggests that "the way members of a group perceive the environment might be a function of the resources and action capabilities that are inherent to the group" (p. 76). Thus, the plank is perceived as being liftable *by us* making the affordance a triadic relation between the plank and two specifically located people. Accordingly, in addition to speaking of affordances for oneself and perceived affordances for another, there may be good reason to speak about the fields of affordances for dyads or groups. This is, we recognize, a somewhat radical proposal. But, as the quote shows, it is already recognized by others, and radical does not mean implausible.

It is clear that the best way to conceptualize the role of affordances in joint action scenarios requires further theoretical work but work that needs to be supported with empirical and formal tools. Accordingly, in Section 4 we argued that our affordance-based account of the mechanisms behind distributive action coordination can be modeled in terms of dynamic field theory. We identified an already realized dyad of cooperating robots as an illustrative case in point. Such an example is not yet available for contributive action coordination. However, we see no obstacles in principle for a DFT model in which action coordination is based on one agent's field of affordances being augmented by perceived affordances for another agent.

We conclude with a preliminary modeling suggestion for formalizing our affordance-based view on contributive action coordination.

In Section 4 we also discussed the application of the DFT framework to perseverative reaching in infants that leads to the famous A-not-B error. It should be noted, however, that the model has changed since its first presentation by Thelen et al. (2001): upon an attempt to replicate the task effects by implementing the model on a simple robot, it has been discovered that instead of moving reliably to either the A or the B location, the robot directions were fluctuating at every time step (Dineva, Faubel, & Schöner, 2007; Spencer et al., 2009). The details of the reason for this problem are not essential for the present discussion, but it is important that the solution to it involved raising the resting level of the motor planning field in the phase in which response was required, i.e. when the containers were moved toward the infant into her peripersonal space. As a consequence, small and unstable bumps of activation could more easily rise above threshold and generate stable reaching decision (Schöner & Dineva, 2007). Given our previous discussion, it should be apparent that a similar process could be taking place whereby the resting level of the movement planning field is affected not just by objects coming into one's own peripersonal space, but also when it is perceived that objects can be acted upon with the other.

Importantly, such a model would capture the recursive-looking bootstrapping of joint affordances. Whether it can be extended to mechanisms spanning multiple individuals and how to best capture such mechanisms with formal tools remains an open question.

7. Conclusions

The aim of this paper was to sketch the outlines of an account of the kind of social cognition involved in simple action coordination that is based on direct social perception (DSP) rather than recursive mindreading. We have not argued directly against a recursive mindreading account. In fact we started out by outlining one mindreading-based account—Michael Tomasello's—as a live option. However, by presenting a DSP account that explains simple action coordination in a less cognitively demanding manner and by arguing that this account allows for formal modeling, part of which has already been accomplished, we have tentatively made an indirect case in favor of DSP in simple action coordination.

Our account of action coordination is based on the notion of an agent's field of affordances. On this account, coordination ensues when, given a shared intention, the actions of and/or affordances for one agent shape the field of affordances for another agent. This a form of *social* perception since in particular perceiving affordances for another person involves seeing that person as an agent. While this does involve attributing some basic kind of intentional relation between a person and her environment, it does not involve attributing propositional attitudes. Affordances were introduced as key ingredients of an ecological theory of perception. Thus, being able to perceive affordances for another person and being able to perceive how another person's actions influence one's own field of affordances counts as a form of social *perception*.

One advantage of our proposal for a DSP account of simple action coordination over other DSP accounts is that it allows for modeling in terms of dynamic fields theory (DFT). We have discussed one existing DFT model that successfully drives and describes action coordination in two robots in a way that neatly fits our proposed DSP account. This example concerns distributive action coordination only. Although a DFT model for contributive action coordination does not yet exist, we argued that there do not seem to be principled obstacles for such a model and made some suggestions for devising one.

References

Bastian, A., Riehle, A., Erlhagen, W., & Schöner, G. (1998). Prior information preshapes the population representation of movement direction in motor cortex. *Neuroreport, 9*(2), 315–319.

Bekkering, H., de Bruijn, E., Cuijpers, R., Newman-Norlund, R., van Schie, H., & Meulenbroek, R. (2009). Joint action: Neurocognitive mechanisms supporting human interaction. *Topics in Cognitive Science, 1*(2), 340–352.

Bicho, E., Hipólito, N., Louro, L., Cambon, S., & Erlhagen, W. (2006). A dynamical field model for joint action in autonomous robots. *Presented at the CogSys II*.

Bohl, V., & Gangopadhyay, N. (2014). Theory of mind and the unobservability of other minds. *Philosophical Explorations, 17*(2), 203–222.

Borghi, A. M., Scorolli, C., Caligiore, D., Baldassarre, G., & Tummolini, L. (2013). The embodied mind extended: Using words as social tools. *Frontiers in Psychology, 4*, 214.

Bratman, M. E. (1993). Shared intention. *Ethics, 104*, 97–113.

Bratman, M. E. (2014). *Shared agency: A planing theory of acting together*. New York: Oxford University Press.

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. Frontiers in Human Neuroscience, 8, 599. Published online August 12, 2014. http://dx.doi.org/10.3389/fnhum.2014.00599 PMCID: PMC4130179.

Cesari, P., Formenti, F., & Olivato, P. (2003). A common perceptual parameter for stair-climbing for children, young and old adults. *Human Movement Science, 22*, 111–124.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15*(2), 181–195.

Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B, 362*(1485), 1585–1599.

Costantini, Marcello, Ambrosini, E., Scorolli, C., & Borghi, A. M. (2011). When objects are close to me: Affordances in the peripersonal space. *Psychonomic Bulletin & Review, 18*(2), 302–308.

Costantini, M, & Sinigaglia, C. (2011). 17 Grasping affordance: A window onto social cognition. *Joint attention: New developments in*.

Dineva, E., Faubel, C., & Schöner, G. (2007). Dynamic field theory and embodiment: Robotic demonstration of infant perseveration. In *14th International conference on perception and*.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91*, 176–180.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review, 109*(3), 545–572.

Eskenazy, T., van der Wel, R. and Sebanz, N. (2012). Mechanisms of skilled joint performance. In N. Hodges and A.M. Williams (Eds.). *Skill acquisition in sports: Research, theory and practice II* (pp. 229–246).

Gallagher, S. (2003). The practice of mind: Theory, simulation or primary interaction. *Journal of Consciousness Studies, 8*(5–7), 83–108.
Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, and Psychology, 11*(3), 199–217.
Gallagher, S. (2007). Simulation trouble. *Social Neuroscience, 2*(3–4), 353–365.
Gallagher, S. (2008). Direct perception in the intersubjective context – social cognition, emotion, and self-consciousness. *Consciousness and Cognition, 17*(2), 535–543.
Gallagher, S., & Zahavi, D. (2008). *The phenomenological mind*. London: Routledge.
Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society, London B, 362*, 659–669.
Gibson, J. J. ([1979] 1986). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
Gilbert, M. (1989). *On social facts*. Princeton: Princeton University Press.
Gilbert, M. (2003). The structure of the social atom: Joint commitment as the foundation of human social behavior. In F. Schmitt (Ed.), *Socializing metaphysics: The nature of social reality* (pp. 39–64). Lanham: Rowman and Littlefield.
Glenberg, A. M., Cowart, M. R., & Kaschak, M. P. (2001). An affordance field for guiding movement and cognition. *The Behavioral and Brain Sciences, 24*(01), 43–44.
Goldman, A. (2006). *Simulating Minds: the philosophy, psychology and neuroscience of mindreading*. New York: Oxford University Press.
Gordon, R. M. (1995). Simulation without introspection or inference from me to you. In M. Davies & T. Stone (Eds.), *Mental simulation: Evaluations and applications*. Oxford: Blackwell.
Herschbach, M. (2008). Folk psychological and phenomenological accounts of social perception. *Philosophical Explorations, 11*(3), 223–236.
Hohwy, J. (2013). *The predictive mind*. New York: Oxford University Press.
Hutto, D. D. (2008). *Folk-psychological narratives: The socio-cultural basis of understanding reasons*. Cambridge, MA: MIT Press.
Hutto, D. and Satne, G. (2015). *The natural origins of content*. Philosophia, forthcoming.
Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism*. Cambridge, MA: MIT Press.
Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport, 7*(14), 2325–2330.
Isenhower, R. W., Marsh, K. L., Carello, C., Baron, R. M., & Richardson, M. J. (2005). The specificity of intrapersonal and interpersonal affordance boundaries: Intrinsic versus absolute metrics. In H. Heft & K. L. Marsh (Eds.), *Studies in perception and action VIII* (pp. 54–58). Mahwah: Erlbaum.
Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science, 20*(5), 568–577.
Kiverstein, J. (2011). Social understanding without mentalizing. *Philosophical Topics, 39*(1), 41–65.
Kopecz, K., & Schöner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields. *Biological Cybernetics, 73*(1), 49–60.
McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review, 90*(3), 215.
McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *PNAS, 98*(20), 11832–11835.
Ratcliffe, M. (2007). *Rethinking commonsense psychology: A critique of folk-psychology*. London: Palgrave Macmillan.
Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology, 26*(4), 325–352.
Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor Cortex and the Recognition of Motor Actions. *Cognitive Brain Research, 3*(2), 131–141.
Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General, 112*(4), 585.
Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
Sandamirskaya, Y, & Schoner, G. (2008). Dynamic field theory and embodied communication. In *Modeling communication with robots and virtual humans* (pp. 260–278).
Sandamirskaya, Yulia, & Schoner, G. (2008). Dynamic field theory of sequential action: A model and its implementation on an embodied agent. In *2008 7th IEEE international conference on development and learning*. http://dx.doi.org/10.1109/devlrn.2008.4640818.
Schöner, G., & Dineva, E. (2007). Dynamic instabilities as mechanisms for emergence. *Developmental Science, 10*(1), 69–74.
Scorolli, C., Daprati, E., Nico, D., and Borghi, A.M. (2011). Catching objects through words. In *17th Meeting of the European society for cognitive psychology*, ESCOP 2011, San Sebastian.
Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
Searle, J. (1995). *The social construction of reality*. New York: Free Press.
Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Science, 10*(2), 70–76.
Slors, M. V. P. (2010). Neural resonance: Neither simulation, nor perception. *Phenomenology and the Cognitive Sciences, 9*(3), 437–458.
Slors, M. (2012). The model-model of the theory-theory. *Inquiry, 55*(5), 521–542.
Spaulding, S. (2010). Embodied cognition and mindreading. *Mind and Language, 25*(1), 119–140.
Spencer, J. P., Perone, S., & Johnson, J. S. (2009). Dynamic field theory and embodied cognitive dynamics. In *Toward a unified theory of development connectionism and dynamic system theory re-consider*. New York: Oxford University Press.
Stoffregen, T. A., Gorday, K. M., Sheng, Y. Y., & Flynn, S. B. (1999). Perceiving affordances for another person's actions. *Journal of Experimental Psychology. Human Perception and Performance, 25*(1), 120–136.
Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *The Behavioral and Brain Sciences, 24*(1), 1–34. discussion 34–86.
Tollefsen, D., & Dale, J. (2012). Naturalizing joint action: A process-based approach. *Philosophical Psychology, 25*(3), 385–407.
Tomasello, M. (2008). *The origins of human communication*. Cambridge MA: MIT Press.
Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origin of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–735.
Tuomela, R. (1992). Group beliefs. *Synthese, 91*, 285–318.
Tuomela, R. (2007). *The philosophy of sociality: The shared point of view*. Oxford: Oxford University Press.
Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. Cambridge MA: MIT Press.
Zahavi, D., & Parnas, J. (2003). Conceptual problems in infantile autism research: Why cognitive science needs phenomenology. *Journal of Consciousness Studies, 10*(9–10), 9–10.