Editorial Manager(tm) for The Review of Philosophy and Psychology
Manuscript Draft

Manuscript Number: ROPP46

Title: Joint Attention: from Interaction to Joint Action

Article Type: Joint Action: What is Shared?

Keywords: joint attention; shared attention, interaction, joint action, intersubjectivity

Corresponding Author: Ms. Anika Fiebich,

Corresponding Author's Institution: Ruhr-University Bochum

First Author: Anika Fiebich

Order of Authors: Anika Fiebich; Shaun Gallagher

Abstract: In this paper, we distinguish between various kinds of interactions and joint actions that emerge on different levels of intersubjectivity during ontogeny. Our main focus is the second level of intersubjectivity around the first year of life. We propose a concept of basic joint action analogous to a concept of basic action and argue that joint attention is the basic joint action, which builds the bridge from interaction to joint action. Finally, we distinguish between joint path-goal actions and joint final-goal actions depending on the nature of the shared goals the agents have.

Suggested Reviewers:

Authors

Anika Fiebich & Shaun Gallagher

Title

Joint Attention: from Interaction to Joint Action

Affiliations and Addresses of the Authors

Anika Fiebich, M.A.
Ruhr-University Bochum
Institute for Philosophy 2
Universitätsstraße 150
44780 Bochum
Germany
Email: anika.fiebich@rub.de
Tel.: 0049-234-32-24725
Fax: 0049-234-32-14963

and

Shaun Gallagher, Prof. Dr.
Philosophy and Cognitive Sciences
Institute for Simulation and Training
University of Central Florida (USA)
School of Humanities
University of Hertfordshire (UK)

Acknowledgments

A.F. acknowledges the support provided by Albert Newen and the Barbara Wengeler Foundation for research stays at the École Normale Supérieure in Lyon and the Institut Jean Nicod in Paris (2010); S. G. acknowledges the support provided by research grants from the Centre National de la Recherche Scientifique as visiting professor at the École Normale Supérieure de Lyon, and visiting researcher at the Centre de Recherche en Epistémelogie Appliquée, École Polytechnique, Paris (2010).

Joint Attention: from Interaction to Joint Action

Abstract

In this paper, we distinguish between various kinds of interactions and joint actions that emerge on different levels of intersubjectivity during ontogeny. Our main focus is the second level of intersubjectivity around the first year of life. We propose a concept of basic joint action analogous to a concept of basic action and argue that joint attention is the basic joint action, which builds the bridge from interaction to joint action. Finally, we distinguish between joint path-goal actions and joint final-goal actions depending on the nature of the shared goals the agents have.

Keywords: joint attention, shared attention, interaction, joint action, intersubjectivity

Children playing tag run down the street towards a young couple who are out for a walk, and they almost run into three men who just come out of a house carrying a piano. Joint actions are frequently encountered in our everyday life. Human beings are essentially social. We live in a complex social system of various roles and rules and communicate by means of a sophisticated symbolic system that is uniquely human; a communication system, which provides the possibility of cultural transfer across generations and has been regarded as the fundamental means by which the enormous cultural and technological development of the *Homo Sapiens* within the last 200,000 years could happen at all (Tomasello 2002). Certain kinds of linguistic actions can be regarded as joint actions (Clark 1996). They can be regarded as constitutive elements for social institutions (Searle 2005), which in turn may pave the way for the regulation and combination of further individual actions into a relatively unified joint action that moves toward a final goal. But what about children in the first year of life who lack linguistic skills and are not aware of all the rules the social system involves? Are they not capable of performing any kind of joint action? In order to answer this question, we clarify what "joint action" actually is and develop a conceptual framework to distinguish different kinds of interactions and joint actions. Then, we propose a concept of basic joint action analogous to a concept of basic actions and argue that joint attention is a basic joint action.

1. Interaction and Joint Action – From the First to the Second Level of Intersubjectivity

Intersubjectivity has been a central topic for philosophers in the phenomenological tradition (see Zahavi 2001 for an overview) as well as for developmental psychologists (Reddy 2008, Trevarthen 1998, etc.). Some of these approaches include the distinction between different kinds of intersubjectivity that develop during ontogeny. Trevarthen (1998), for example, distinguishes between *primary intersubjectivity* and *secondary intersubjectivity.* Human beings experience *primary intersubjectivity* from birth onwards, characterized by infants' sensitivity and emotional regulation of the others' bodily gestures and facial expressions perceived in dyadic interactions. *Secondary intersubjectivity* emerges around one year of age and is defined by the "systematic combining of purposes directed to objects with those that invoked interest and interaction from a companion" (Trevarthen 1998, p. 31). Using an alternative set of distinctions, Fiebich, deBruin and Newen (2010, submitted) have argued that human beings interact on three different levels of intersubjectivity according to a growing set

of cognitive competencies involved in the social understanding process. This model does not contradict Trevarthen's, but it has some advantages over his distinction since, based on current studies from developmental psychology, it provides a careful analysis of different kinds of self-other distinctions and different kinds of perspective-taking involved in the social understanding process.

The three-level model of intersubjectivity includes

- Level 1: which involves a *bodily* self-other distinction and an immediate sense of the other person as an agent;
- Level 2: which involves an *emotional* self-other distinction and an understanding of the other as someone having their own perceptual-attentive perspective and affective attitudes towards external entities (such as persons, objects and events); and
- Level 3: which includes a *conceptual* self-other distinction and an understanding of the other person as someone having their own intellectual perspective, i.e., their own beliefs and desires towards external entities.

In the present paper, we focus on the second level of intersubjectivity and argue that joint attention is a basic joint action that builds the bridge from dyadic social interactions on the first level of intersubjectivity to joint actions involving triadic attentional structures. Joint attention first occurs on the second level of intersubjectivity and forms the foundation for more sophisticated kinds of joint action that develop later on.

Human beings are engaged in social interaction from birth on. Even newborns start to interact with their social interaction partner on a first level of intersubjectivity in an embodied and stimulus-dependent way as a number of studies conducted in the late 1970s on neonate imitation show (Meltzoff and Moore 1977, Field et. al. 1982, etc.). Although it might be debatable whether newborn imitation reflects a voluntary performed action by the infant or is based on perceptual priming or contagion (see Hurley and Chater 2005; Meltzoff and Prinz 2002), other studies present evidence that newborns sense others as agents different from themselves since they are able to distinguish their own actions from those of someone else. Rochat and Hespos (1997), for example, showed that newborns react with significantly more rooting responses (i.e. head-turning towards the stimulation with open mouth) if the experimenter touches one of the infant's cheeks (external stimulation) compared to if the infant herself touches her own cheek (self-stimulation).

We account for social understanding on different levels of intersubjectivity in terms of an embodied "Interaction Theory" of social cognition (Gallagher 2001; 2004; 2008). In the first weeks of post-natal life infants are sensitive to the intentions and emotions of their social interaction partner, which are embodied in the other's bodily movements, gestures, facial expressions, tone of the voice, etc (Baldwin et al. 2001; Hobson 2004; Trevarthen 1979). Infants coordinate their responses in time-sensitive and emotion-sensitive interactive processes and they show patterns of distress if the social interaction partner reacts inappropriately to their actions (Murray and Trevarthen 1985, Nagy 2008). These embodied practices continue to provide our primary access for understanding others throughout our life span (Gallagher 2008). They correspond to the first level of intersubjectivity (indicated above), as well as to Trevarthen's concept of primary intersubjectivity.

Although there is evidence for a bodily self-other differentiation at the earliest age (at Level 1), there is no evidence for an emotional self-other differentiation. Rather, the infant experiences the other in a unified emotional experience of interaction. Even newborns

younger than 4 days old show patterns of distress if this emotional connection is disrupted and if their social interaction partner meets them with a neutral instead of emotional and responsive facial expression (Nagy 2008).

From the 3rd/4th month on, a second level of intersubjectivity emerges. Infants start to differentiate others emotionally. Rather than reacting with a smile to each person who approaches them in a friendly manner, they show patterns of anxiety towards strangers (Sroufe 1977, Sokolowski and Heckhausen 2006). The motivation of newborns to share emotions with others becomes person-dependent from this age onward, and the experience of the emotions of the other seems to involve a simple emotional self-other distinction. Although the infant may recognize the happiness that the other expresses by smiling to her, the infant herself may feel afraid of the other and start crying (instead of smiling back).

Additionally, social understanding on the second level of intersubjectivity involves an, at least implicit, comprehension of the other as having its own perceptual-attentive perspective. This is indicated by 3-month-olds who start to follow the gaze (or first the head movement[1]) of the other towards a perceivable entity (i.e., another person, an object or an event) in their (the infant's) field of vision, i.e., if the entity is spatially close enough to the other (D'Entremont 2000). If one assumes, following Tomasello et al. (2005), that infants are motivated to share the attention, emotions, desires and beliefs of others as soon as they understand them in others, then following the gaze or head movement of the other might indicate an (at least implicit) understanding of the perceptual-attentive perspective of the other. In general, it is an open question whether in gaze-following the child sees the object to which the other is attentive, or just looks in the direction in which the other is looking (Baron-Cohen 1995).

In the present paper, we focus on the second level of intersubjectivity. We look at evidence that shows that 3-month-old infants start to go beyond simple dyadic person-person interactions (in which they are engaged on the first level of intersubjectivity). Part of what allows for this are developing capabilities for entering into different forms of intersubjective attention. On our view, this development culminates with joint attention which enables the sophisticated form of interaction called joint action.

2. From Simple to Joint Attention

From birth on, infants are attentive towards external entities (call this 'simple attention') and are engaged in dyadic self-other interactions which involve dyadic attention where subjects are mutually attending to each other (see Table 1). From the 3rd month of life onward, a new level of intersubjectivity emerges that involves, first, the capacity of recognizing the perceptual-attentive perspective of the other, and second, not just a bodily but also an emotional self-other differentiation (Fiebich et. al. submitted). As mentioned above, infants at this age start to follow the gaze of the other. Interestingly however, 3 month-olds, in contrast to 6-9 month-olds, don't follow the gaze of the other if the other shows facial expressions of joy or disgust (De Groote et. al. 2007). These findings might be interpreted in various ways, e.g., that the child tends to understand each emotional expression of the other as related to herself (without understanding the link between gaze direction and emotional expression), or that the emotional facial expression is simply so fascinating to the child that she keeps staring

[1] Meltzoff and Brooks (2007) show that infants younger than 9 months follow head movement regardless of whether eyes are open or closed. Thus, they follow the head rather than the gaze. With this in mind, we use the term 'gaze-following' broadly to include following head as well as eyes.

at the other. Regardless of how these findings are interpreted, Moore (2008, p. 68) suggests, "it appears that for younger infants, emotional expression tends to elicit dyadic rather than triadic forms of interaction – that is, any emotion focuses the infant's attention on the adult instead of the adult's gaze towards the target." Moore assumes here that when gaze-following does happen, it elicits a triadic form of interaction, because the infant is aware of the adult being attentive towards the object *and* herself being attentive towards the object. We propose to call this 'shared attention', where the infant (or any agent) is aware of herself and the adult (or the other) being attentive towards the same object. However, shared attention, on this definition, is not yet a triadic form of interaction; it is rather a dual attending – a combination of simple attendings: a simple attending to the object, and a simple attending to the person being attentive towards the same object. It might also be the case that the child, once she follows the gaze of the adult forgets that the adult is attending to the object (maybe because of limited memory skills or because the object is so fascinating that it causes her to forget everything else in the environment including the adult). In that case there is shift from a simple attention (directed toward the adult) to a simple attention (directed toward the object).

From 9-10 months onward, in the phenomenon of social referencing, for example, when infants start to refer to the other's emotional expression to know whether to approach novel objects (Klinnert et. al. 1986, Moses et. al. 2001), the infant is engaged not just in shared attention but in *mutual* shared attention, i.e., an interaction that includes the awareness that she and the other are both attentive towards the same entity *and* that the other is aware of this shared attention as well. Infants learn that monitoring the gaze direction of the other enables them to predict the location of interesting objects in the environment (Moore and Corkum 1994). In line with this research, Striano and Rochat (2000) showed that whether 10-month-olds (but not 7 month-olds) monitor and refer to others in ambiguous situations depends on the others' attention towards them (i.e., the infants). In the famous "visual cliff experiment" conducted in the 1960s, Gibson and Walk showed that the child refers to the mother in order to check whether it is safe to scramble over a visual cliff (i.e. a see-through Plexiglas surface) or not. Infants cross the cliff only if the mother shows positive facial expressions.[2] In social referencing, the child aims for mutual shared attention towards the entity for certain purposes (e.g. to find out whether it is safe to approach the novel entity or not). The mother might grasp this individual intention and respond to it, but, as we'll see, if there is no *shared* intention involved here, it is not yet a joint action.

Mutual shared attention might also be elicited because of the ostensive cues of the other; if the other looks at the infant or waves to her before he starts dealing with an object, the infant is aware of herself and the other as being attentive towards this object and the other as being aware of that as well – this mutual shared attention plays a crucial role in social learning of object features. Yoon, Johnson and Csibra (2008) showed e.g., that if 9-month-olds perceive an object as the referent of ostensive communication (and thus are engaged in a triadic interaction involving mutual shared attention), they remember better its permanent, kind-relevant properties (e.g. its shape) whereas they remember better the object's transient properties (e.g. its location) if their attention is drawn to the same object by non-communicative means (i.e., if they are engaged in shared attention). We note that in cases of shared attention in adults, object evaluation is influenced by less ostensive, more implicit facial expressions of the other. Bayliss et al. (2006) showed that subjects evaluate objects as more (or less) likable if the other looked towards (or away) from that object and that this effect was even stronger if an emotional expression was added to the other's face. These

[2] Recent studies suggest, however, that it may be the positive vocal cues rather than the visual cues of the mother that determines the infant's behavior (Vaish and Striano 2004).

studies suggest that in shared attention in adults, the recognition of the affective attitude of the other towards an object plays a crucial role in the subsequent object evaluation. Developmentally, this type of effect may begin with social referencing at around 10 months. The impact of ostensive cues or the lack of them on social learning led Csibra (2010, p. 156) to the following view:

> [Infants] are biased to assume generalizability of the content of the informative intention signified by the ostensive signals. […] When they observe the emotional expressions in a non-communicative context, they do not think that the actor's personal preference extends to other people. In contrast, when the emotional expressions are presented for them ostensively, they apparently think that the same preference applies to others as well. Thus, they generalize the object valence information acquired from a single person to every, one, or, in other words, they treat this information as the property of the object and/or its kind rather than of the actor.

To summarize thus far, on the second level of intersubjectivity, between 3 and 10 months of age, infants enter into relations characterized by shared attention, and then mutual shared attention. In these relations the recognition of emotional expression by the other plays a crucial role for social learning. Mutual shared attention, however, is not yet joint attention, which requires both that the shared attention be mutual, and, on our view, that there be a shared goal.

It is important to note that the terms 'shared attention' and 'joint attention' are defined in various ways in the current literature, and are sometimes used interchangeably (see Triesch et al. 2006). According to Baldwin (1995) e.g., joint attention simply means "the simultaneous engagement of two or more individuals in mental focus on one and the same external thing" (p. 132); this is what we define as 'shared attention', but only when one of the individuals is aware of himself and the other as being attentive towards the same external entity. According to other theorists (e.g. Tomasello et al. 2005) joint attention includes the mutual awareness of both individuals as being attentive towards the same external entity; this is what we call 'mutual shared attention'. On our view, what needs to be added to mutual shared attention to make it 'joint attention' is a shared intention. That is, both individuals are not just mutually aware of being attentive towards the same entity, but also *intend* to be mutually attentive as a shared intention. Furthermore, whereas in mutual shared attention both agents are attentive towards the same entity in the environment, in joint attention they are (also) mutually attentive towards the shared goal or mutually share the same purpose. Mutual shared attention itself can (but does not need to) elicit an immediate shared intention, because of the environmental circumstances or the intentional behavior of one of the agents. If it does, however, it becomes joint attention (see Table 2).

3. Joint attention as basic joint action

Interaction is a kind of relation between two (or more) individuals who mutually influence one another (DeJaegher, DiPaolo, and Gallagher 2010). As we have discussed above, not only mothers but also their newborns respond sensitively to the interactive behavior patterns of the other (Nagy 2008) on a first level of intersubjectivity. Joint action, however, is a form of interaction that involves more than interaction. There are different approaches in the current literature, however, to what exactly joint action involves. Sebanz, Bekkering, and Knoblich (2006) propose joint action as "any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (p. 70). They discuss joint attention, action observation, task-sharing and action coordination as cognitive mechanisms through which a successful joint action can be achieved. Tomasello and colleagues (2005) would add a shared goal to this definition of joint

action and argue for the shared goal as a crucial part of what distinguishes human group activities from those of other animals. Carpenter (2009) highlights that "shared intentionality is the foundation upon which joint action is built" (p. 381). Along the same line, Searle (1990) defines joint action as collective intentional behavior involving We-intentions. Collective intentional behavior is not the same as the summation of individual intentional behavior; thus, we-intentions cannot be analyzed into sets of I-intentions, even if these I-intentions are supplemented with beliefs about the intentions of others. Bratman (1992) on the contrary indicated three characteristic features of joint action: (1) mutual responsiveness to the intentions and actions of the other, (2) commitment to the joint activity, and (3) commitment to mutual support of the efforts of the other to play her role in the joint activity. In joint action "each agent does not just intend that the group perform the (cooperatively neutral) joint action. Rather, each agent intends as well that the group perform this joint action in accordance with subplans (or the intentions in favor of the joint action) that mesh" (Bratman 1992, p. 332). These intentions need to be common knowledge between the agents.

In line with this current literature, we define joint action of a group of agents as involving three conditions satisfied by the agents:

(1) a shared goal/shared intention

(2) common knowledge of aiming for this goal

(3) participation in coordinated behavior patterns (determined by rules or regularities) in order to achieve this goal [and to mesh sub-goals]

Whether and to what extent commitments are involved, or whether meshing sub-goals is required, depends on the shared intention or goal of the particular joint action. One can distinguish between two kinds of joint action; (1) *joint path-goal actions* (in which the shared goal entails only the activity itself, and thus, the coordinated behavior patterns to achieve the shared goal are specified in advance), and (2) *joint final-goal actions* (in which both agents aim for a certain end product or state, which might be independent from a specific coordinated behavior pattern to achieve that final goal). The more rule-governed the shared goal (i.e. the more the coordinated behavior patterns are specified in advance, which is very much the case in *joint* path-goal actions), the more commitments are involved and the higher are the action expectations, but the less meshing sub-goals is necessary. For example, if we decide to play chess together, we commit to observe the rules entailed in chess and our action expectations in playing chess are based on these commitments; thus, you would protest if I used the rook like the queen. But once we accept all the rules entailed in chess, no sub-goals are left that might diverge and stop our joint activity since our action opportunities within the game are prescribed in advance[3]. Additionally, we not only committed to observe the rules of chess when we decided to play chess together but also to finishing the play. Thus, if your brother, who is an excellent chess player and with whom playing chess might be much more exciting for you than playing with me, arrives, I nonetheless expect you to finish the game with me.

On the contrary, if we decide to cook together (where we aim for a shared final goal, which is to have dinner as the end product of our cooking), there is still a lot of space left in this shared goal for diverging sub-goals; e.g., you might think of a rare steak whereas I prefer a vegan dish. Even if we both decide to cook a well-known dish like Spaghetti Bolognese there might

[3] There is no space left for sub-goals concerning the joint activity itself but there may be diverging sub-goals concerning the situational features (location, date, etc.) the joint activity takes place in.

be divergences in how we go about preparing the dish – I might dislike onions whereas you don't want to leave this ingredient out of the sauce, which might mean ending up not cooking Spaghetti Bolognese together. But (at least) one of us could agree to deviate from her original sub-goal so that we don't give up our shared goal. In other cases we might have different sub-goals that do not contradict each other, e.g., you might have the sub-goal to eat Spaghetti Bolognese with onions whereas I have the sub-goal that we eat at my home. But also joint final-goal actions might (but do not need to) specify coordinated behavior patterns to achieve that goal in advance, which is the case if we decide to cook a specific recipe.

Although we take rule-governed (conventional) commitments and meshing sub-goals to be important parts of some joint actions, we do not take them as necessary parts (and thus we bracket these in the above definition) since in some kinds of joint action the coordinated behavioral patterns are so minimal or so rule-governed that they do not leave any space for sub-goals concerning the activity itself. Setting aside these bracketed issues, then, on this definition of joint action we propose that joint attention is a basic joint action, which builds the bridge from interaction to joint action.

On the second level of intersubjectivity, the 3-month-old shows a first implicit understanding of the perceptual-attentive perspective of the other, which is indicated by the phenomenon of gaze-following (De Groote et. al. 2007). If the child is simultaneously aware of herself being attentive towards an external entity and the other being attentive towards the same entity they are engaged in shared attention. Even if this becomes mutual shared attention with the addition of mutual awareness, it is not yet joint action because it doesn't meet the three conditions just defined. In the case of mutual shared attention, there is no shared goal (condition 1) involved, and thus, neither a common knowledge for aiming for this goal (condition 2) nor participation in a coordinated behavior pattern to achieve this goal (condition 3).

In mutual shared attention the infant is aware of herself and the other being attentive towards an external entity and aware of the other being aware of that as well. When mutual shared attention is involved, both agents interact in a triadic relation (agent-agent-object). Mutual shared attention itself may (but does not necessarily) elicit a shared intention-in-action, maybe because of the environmental circumstances or because of the intentional behavior of one of the agents. If it does, it becomes joint attention and all three conditions of a joint action are fulfilled. There is at least a minimal shared goal (condition 1) of maintaining joint attention (e.g., when we are both surprised by some object or event and desire to jointly remain attentive, perhaps for the sake of mutual enjoyment). There may be other goals involved as when jointly attending to some task. Even in the minimal case there is common knowledge (and likely desire) to maintain joint attention as the goal (condition 2), and to coordinate our behavior patterns to achieve this goal (condition 3).

Given our definition of joint action, we argue that joint attention is a *basic* joint action on analogy with what theorists such as Danto (1973), Davidson (2001), and Hornsby (1980) call *basic actions*. Basic actions are identified as unmediated bodily movements that serve other actions which have a higher-order description. In the case of a normal, complex, intentional action, if my intention is to get a drink, this is made up of more basic actions such as walking across the room, reaching and grasping the drink, etc. Not only are these basic actions intentional movements on my part, but without these basic actions I would never be able to get my drink. Likewise, we suggest, joint attention is the basic joint action that plays a role in all joint action. In all cases, joint attention serves purposes other than itself and is rather meaningless (an artificial abstraction) if it doesn't. Even in the minimal case, where our goal

is to maintain joint attention, it is for the sake of something, e.g., enjoyment. Joint attention, as basic joint action, derives its meaning from the shared intentions that it serves.

The concept of basic action, however, is a controversial one in action theory (see e.g., Dennett, 1984, p. 111; Sneddon 2006, pp. 98ff). Sneddon considers basic actions to be theoretical entities the existence of which requires demonstration (2006, p. 101). The usual strategy in this regard is (1) to point to the phenomenology of action – i.e., that some actions seem to be unmediated and directly willed – or (2) to offer an argument involving avoidance of infinite regress – i.e., that not all actions can be mediated (see. E.g., Danto, 1979, p. 46). Sneddon, however, argues that basic actions are mere abstractions from normal complex actions, rather than real actions themselves. On this view, the phenomenology, rather than supporting the existence of basic actions, is rather a reflective abstraction that produces a theoretical analysis of action. Furthermore, the avoidance of infinite regress means that at the bottom level we find processes that do not count as genuine actions, i.e., bodily movements should not be considered actions *per se*. Here one might accept that basic actions are only actions in a derivative sense since their intentional status and important aspects of their motor control derive from the complex intentional actions that they serve.

The controversies surrounding basic actions, however, do not apply to the concept of joint attention understood as basic *joint* action since whereas the debate around basic actions concerns actions of a single individual, thus involving *individual* intentions, basic joint actions involve (at least) two agents and their *shared* intentions. Joint attention is therefore not an abstraction – it's a real interactional process in which we engage with others. Indeed, one can analyze joint attention into more subtle components of interaction (those that involve mutual shared attention or the various bodily movements that constitute the required coordination). But, as we have shown, such components should not be considered more basic joint actions. Accordingly, there is no infinite regress to avoid. What turns out to be more basic turns out not to be more basic joint actions, but merely basic actions (or bodily movements of coordination). Since joint attention, insofar as it involves (1) a shared goal, (2) common knowledge of intending this goal, and (3) participation in coordinated behavior patterns in order to achieve this goal, is a joint action, and, on analysis, its components are not themselves basic joint actions, then it is a basic joint action.

4. Shared Intentions in Joint Actions

Joint actions are frequent in our everyday life. "Shared cooperative activities can involve large numbers of participating agents and can take place within a complex institutional framework – consider the activities of a symphony orchestra following its conductor" (Bratman 1992, p. 327). Keeping things simple, we consider in the present paper joint actions that involve just a pair of participating agents at the second level of intersubjectivity. First, however, we give a general introduction into different kinds of joint actions that we distinguish depending on which kind of shared intention they involve.

All joint actions involve "WE-Intentions" that cannot be analyzed into sets of I-intentions and cannot be expressed without reference to the other, although the specific bodily movements of collective intentional behavior may be indistinguishable from individual intentional behavior (Searle 1990). Some philosophers offer an intention theory of action to explain individual intentional behavior (Searle 1980, Bratman 1987, Pacherie 2008). For our purposes, we use Searle's conceptual framework, not to explain individual actions, but to explain the collective intentional behavior found in joint actions.

For Searle (1980) prior intention means "the agent has the intention to perform the action prior to the performance of the action itself" (p. 52). In contrast, actions that involve intentions-in-action are spontaneously performed and do not involve any prior intentions. But in order to achieve a prior intention, such as driving to my office, several intentions-in-action may be performed, such as opening the car door, starting the engine, depressing the clutch, etc. "For such cases, I have an intention, but no prior intention" (Searle 1980, p. 52).

We acknowledge the importance of this conceptual difference not just for individual intentional behavior but also for collective intentional behavior. Different kinds of joint actions need to be distinguished according to different requirements for shared prior intentions (or shared goals). This is the case because, according to our definition of joint action, both agents need to have a shared goal, be mutually aware of aiming for that goal, and coordinate their behavior in order to achieve that goal. In some situations where the setting of the shared goal and the common knowledge of aiming for that goal are not prior to the activity itself, it would be paradoxical if the agents were able to coordinate their behavior to achieve that shared goal. In other situations, however, prior shared intentions are not required. For example, if someone simply grabs another person and starts to dance, the shared goal – of dancing – only emerges in the already ongoing action. The other person says, after the fact "Ok, let's dance". Only at that point does the shared goal emerge. Coordinated behavior patterns themselves however involve shared intentions-in-action. In some situations of joint action one requires shared intentions-in-actions for coordinated behavior patterns performed in order to achieve a shared prior intention.

To clarify these adumbrations in regard to the relation of shared intentions and joint actions, we return to the distinction between two different kinds of joint actions, which depends on the shared goal/intention involved: (1) *joint path-goal action*, and (2) *joint final-goal action*. Based on the above analysis we suggested that complex joint actions include joint attention as a basic joint action. Here, however, we need to distinguish between two different kinds of joint attention: (a) *goal-directed joint attention* (where what we jointly attend to is the goal of the joint action) and (b) *action-oriented joint attention* (where we jointly attend to the action itself, which, in the case of joint path-goal action just is the goal). Whereas all joint actions include goal-directed joint attention and all joint path-goal actions include action-oriented joint attention, action-oriented joint attention may but does not need to be included in joint final-goal actions.

In some *joint path-goal actions*, the coordinated behavior patterns required to achieve that goal are determined by the shared prior intention to such a high degree that the shared goal just is the action as an end in itself. If we decide to go for a walk together (which is our shared prior intention), the coordinated behavior patterns to achieve this goal (walking together) and the goal itself are the same. Furthermore, we aim not just for joint attention towards the shared goal (i.e., goal-directed joint attention) but also for joint attention towards the activity itself (i.e., action-oriented joint attention). Without the latter a coordinated behavior, which is required to achieve the goal, wouldn't be possible. It is not joint action if both agents simply coordinate their behavior without a shared intention (e.g. if a child simply follows her mother). In more complex forms of joint path-goal actions, conventional commitments may be included, e.g., in playing a conventional game like chess. But also here, the shared goal (to play chess) and the coordinated behavior patterns to achieve this goal (playing chess) are the same. Whether there is a final end product or state included in the shared goal or not (me winning the game is not a shared goal, but that someone wins may be part of the shared goal), it is worth mentioning that the shared goal includes performing the action *to its end*, i.e., the commitment to finish the joint action regardless any sub-goals.

In *final-goal joint actions*, the shared prior intention entails a final end-product or state. If we decide to do a puzzle together, we aim for a certain final product of our joint activity, namely for creating a completed puzzle. Even in such a final-goal joint action, we might like the coordinated behavioral patterns that are required for achieving the final product so much that we may partly have as our shared goal doing the puzzle because of the activity itself. But we could also dislike the activity itself and aim nonetheless for the shared goal to get the puzzle done. Final-goal actions are independent of whether the shared goal involves the sub-goal of performing the activity itself. Rather they are defined by the final product aimed for in the goal-directed joint attention.

In some final-goal joint actions coordinated behavior patterns are required in order to achieve the goal, and if so, action-oriented joint attention is included; e.g., if we cook together and you see me salt the soup, you wouldn't salt the soup again. In other final-goal joint actions however, no action-oriented joint attention is necessary. Imagine two terrorists aiming for the shared goal of blowing up Big Ben, each one knowing that they share the same goal, and each one knowing that the goal will be reached only in coordinated actions. Neither terrorist, however, knows precisely what the other one will be doing. They wait at different places for instructions from their controller, and their behaviors are coordinated by his instructions. Here, no action-oriented joint attention is involved; since one does not interact directly with the other and each is located in a different place, each is attentive only towards his own activity and no mutual shared attention towards the other's activity is possible. Nonetheless, all three defining conditions of joint action are fulfilled, and the joint attention involved is strictly goal-directed joint attention. In effect, joint final-goal actions always involve goal-directed joint attention, but not necessarily action-oriented joint attention.

Joint actions can be planned in advance, and if so, they often involve certain spatiotemporal coordinates – when and where the action is to be performed. But joint actions can also emerge spontaneously (as in the dance example) and from mutual shared attention due to the environmental circumstances of the social interaction or the intentional behavior of one of the agents. Imagine we are sitting on a bench at the edge of a field and are suddenly mutually aware of a herd of cattle stampeding towards us. This leads us immediately to form the shared intention of making a run for it; we both know how to do this and we coordinate our behavior in order to achieve that goal by grabbing our jackets and stuff, getting up and starting to run. Mutual shared attention itself may be elicited by the intentional behavior of one of the agents (which transforms a dyadic relation into a triadic one – e.g., I see the cattle coming and then alert you), which may in turn elicit an immediate joint action, lacking anything like a prior intention.

Or imagine another example. We are sitting together in my living room and I get up, walk to the bookcase and reach for a book on the top shelf; because I am too short, I am unable to reach it. You, as short as I am, follow me and pick me up so that I can reach it. First, my intentional behavior elicits your attention towards the book and since you are still in my field of vision while I reach for the book, we are mutually aware of being attentive towards the same book. Second, my intentional behavior brings you to adopt my individual goal to get the book (which transforms it into a shared goal), and to help me to achieve that goal in a coordinated behavioral way. If you do not adopt my individual goal and help me (for whatever reason), it involves mutual shared attention but not joint attention. Likewise you may not form the intention to run away when you see the cattle (maybe because you notice a fence that will prevent the cattle trampling us).

What we want to draw out of these various examples, is that in joint actions there are matters of degree and a variety of ambiguities concerning the formation of shared intentions, the role

of different forms of joint attention, and the type of joint action involved. Despite this, we think we can make some headway in mapping out these various factors and in collating this map with developmental studies.

Infants are engaged in joint actions as soon as they are able to grasp that the other's intention is the same as their own, and to coordinate their behavior patterns with those of the other. However, before they can grasp the intention of the other as shared, they need to be able to understand the intentional behavior of the other *simpliciter*. Philipps, Wellman, and Spelke (2002) showed that infants' ability to connect gaze and emotional expression to intentional action emerges at the end of the first year of life; whereas 12-month-olds recognize that an actor is likely to grasp one object (among others) which he had visually regarded with positive emotional expressions, 8-month-olds don't. A number of studies have shown that infants from 1 year of age onward are not only capable of grasping the intentional behavior of the other but may also engage in helping the other to achieve his goal (i.e., they are engaged in joint final-goal actions), for example, by pointing to the needed information (Liszkowski et. al. 2006) as well as more instrumentally by e.g., opening cabinet doors for them (Warneken and Tomasello 2006). In task-sharing situations, even 18-month-olds show that they understand the commitment to finish the task by waiting for the adult or trying to reengage him if he suddenly stops the joint action (Warneken, Chen, and Tomasello 2006, see also Carpenter 2009). Since infants were enthusiastically engaged in activities that entailed no material reward, "the collaborative activity was thus an end in itself, not just a means to retrieving some material reward" (Carpenter 2009, p. 386), i.e., they were engaged in joint path-goal action.

5. Summary and Outlook

On the second level of intersubjectivity, infants go beyond dyadic person-person interactions in which they are already engaged on the first level, and enter interactions that involve shared attention and mutual shared attention. Joint attention builds a bridge from dyadic interactions to joint actions, and is itself a basic joint action. We distinguish between joint path-goal actions and joint final-goal actions depending on the nature of the shared goals the agents have. These distinctions can be mapped onto a developmental account supported by recent empirical literature. Although we have focused on the second level of intersubjectivity, there is much more to be said about joint action on the third level of intersubjectivity. Development of language and communicative skills and more sophisticated understanding of social roles and rules allow children to enter into complex joint actions that take place in the social, cultural and institutional systems we live in.

References

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, 72 (3), 708-17.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origin and Role in Development* (pp. 131-159). Hillsdale, N. J.: Lawrence Erlbaum.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Bayliss, A.P., Paul, M.A., Cannon, P.R. & Tipper, S.P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychological Bulletin & Review*, 13(6), 1061-1066.

Bratman, M.E. (1992). Shared Cooperative Activity. *The Philosophical Review*, 101(2), 327-341.

Bratman, M.E. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Cambridge University Press.

Carpenter, M. (2009). Just how joint is joint action in infancy? *Topics in Cognitive Science*, 1, 380-392.

Clark, H.H. (1996). *Using language*. Cambridge. England: Cambridge University Press.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, 25, 141-168.

Danto, A. (1973). *Analytic Philosophy of Action*. Cambridge: Cambridge University Press.

Danto, A. (1979). Basic actions and basic concepts. *Review of Metaphysics* 32 (3): 471-85.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.

De Groote, I., Roeyers, H., & Striano, T. (2007). Gaze following as a function of affective expression in 3-, 6-, and 9-month-old infants. *Infant Behavior and Development*, 30, 492–498

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Does social interaction constitute social cognition? *Trends in Cognitive Sciences*. Published online 30 July 2010. 10.1016/j.tics.2010.06.009

Dennett, D. (1984). *Elbow Room*. Cambridge, MA: MIT Press.

D'Entremont, B. (2000). A perceptual-attention explanation of gaze-following in 3- and 6-month-olds. *Developmental Science*, 3, 302–311.

Fiebich, A., deBruin, L. & Newen, A. (2010). *Three Levels of Intersubjectivity*. Paper presented at the Conference on "Intersubjectivity and the Self", Budapest.

Field, T.M., Woodson, R., & Greenberg, R. (1982). Discrimination and imitation of facial expressions by neonates. *Science*, 218, 179-181.

Gallagher, S. (2001). The practice of mind: Theory, simulation, or interaction? *Journal of Consciousness Studies*, 8(5–7), 83–107

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. (2008). Inference or interaction: Social cognition without precursors. *Philosophical Explorations*, 11(3), 163-73.

Gibson, E.J. & Walk, R.D. (1960). The "visual cliff". *Scientific American*, 202, 64-7.

Hobson, P. (2004). *The Cradle of Thought*. London: Macmillan.

Hornsby, J. (1980). *Actions*. London: Routledge and Kegan Paul.

Hurley, S. and Chater, N. (eds.) (2005). *Perspectives on Imitation: From Neuroscience to Social Science* (2 vols). Cambridge, MA: MIT Press.

Klinnert, M.D., Emde, R.N. Butterfield, P., & Campos, J. J. (1986). Social referencing: The infant's use of emotional signals from a friendly adult with mother present. *Developmental Psychology* 22 (4): 427-432

Liszkowski, U., Carpenter, M., Striano, T. & Tomasello, M. (2006). Twelve- and 18-month-olds point to provide information for others. *Journal of Cognition and Development*, 7, 173-187.

Meltzoff, A. N., & Brooks, R. (2007). Eyes wide shut: The importance of eyes in infant gaze following and understanding other minds. In R. Flom, K. Lee, & D. Muir (Eds.), *Gaze following: Its development and significance* (pp.271-241). Mahwah, NJ: Erlbaum.

Meltzoff, A. N., & Prinz, W. (2002). *The imitative mind: Development, evolution, and brain bases.* Cambridge: Cambridge University Press.

Meltzoff, A.N. and Moore, *M.K. (*1977*).* Imitation of Facial and Manual Gestures by Human Neonates, *Science,* 198, 75-78.

Moore, C. (2008). The Development of Gaze-Following. *Child Development Perspectives*, 2(2), 66-70.

Moore, C. & Corkum, V. (1994). Social understanding at the end of the first year of life. *Developmental Review*, 14, 349-372.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, 72, 718–735.

Nagy, E. (2008). Innate intersubjectivity: Newborn's sensitivity to communication disturbance. *Developmental Psychology*, 44(6), 1779-1784.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, 107, 179-217.

Philipps, A.T., Wellman, H.M., Spelke, E.S. (2001). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, 85, 53-78.

Reddy, V. (2008). *How Infants Know Minds*. Harvard: Harvard University Press.

Rochat, P. & Hespos, S.J. (1997). Differential Rooting Responses by Neonates: Evidence for an Early Sense of Self. *Early Development and Parenting*, 6, 105-122.

Searle, J. (1990). Collective Intentions and Actions. In Cohen, P., Morgan, J. and Pollack, M. (Eds.), *Intentions in Communication* (pp. 401-415). Cambridge, MA: MIT Press.

Searle, J. (2005). What is an institution? *Journal of Institutional Economics*, 1(1), 1-22.

Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Searle, J. (1980). The Intentionality of Intention and Action. *Cognitive Sciences*, 4, 47-70.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70-76.

Sneddon, A. (2006). *Action and Responsibility*. Dordrect: Springer.

Sokolowski, K. & Heckhausen, H. (2006). Soziale Bindung: Anschlussmotivation und Intimitätsmotivation. In Heckhausen, J. & Heckhausen, H. (Eds.), *Motivation und Handeln* (pp. 193-210). Heidelberg: Springer.

Srouffre, L.A. (1977). Wariness of strangers and the study of infant development, *Child Development*, 48, 731-746.

Striano, T. & Rochat, P. (2000). Emergence of Selective Social Referencing in Infancy. *Infancy*, (2), 253-264.

Tomasello, M. (2002). *Die kulturelle Entwicklung des menschlichen Denkens*. Frankfurt a.M.: Suhrkamp.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675-735.

Trevarthen, C. (1998). The concept and foundation of infant intersubjectivity. In S. Braten (Ed.), *Intersubjective Communication and Emotion in Early Ontogeny* (pp. 15-46). Cambridge: Cambridge University Press.

Triesch, J., Teuscher, C., Deák, G., & Carlson, E. (2006). Gaze following: why (not) learn it? *Developmental Science*, 9(2), 125-147.

Vaish, A. & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior, *Developmental Science*, 7(3), 261-269.

Warneken, F. & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 3, 1301-1303.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development*, 77, 640-663.

Yoon, J. M. D., Johnson, M. H. & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 13690–5.

Zahavi, D. (2001). Beyond Empathy. Phenomenological Approaches to Intersubjectivity. *Journal of Consciousness Studies*, 8(5-7), 151-167.

Table 1: Different forms of attention

Simple attention: the person attends to/observes an entity, i.e. an object, event or other person (but without the other person attending back)
Dyadic attention: the person attends to another agent in an interactive relation, i.e., with the other person attending back
Shared attention: a combination of simple attention to an object, and simple attention to the other person being attentive towards the same object (e.g., gaze following)[1]
Mutual shared attention: a triadic combination of simple attention to an object and dyadic (mutual) attention to the other person as being attentive towards the same object (e.g., interactional gaze following)
Joint attention: mutual shared attention with a mutually acknowledged purpose

[1] This may be purely observational rather than interactional – i.e., without the other person attending back. Or it may happen in cases where the two people are interacting but one is not aware that the other is being attentive towards the same object.

Table

Click here to download Table: TABLE 2 joint attention - from interaction to joint action.doc

Table 2: From Simple to Joint Attention

	Attending to one entity	Attending to another person in an interactive relation	Dual simple attending (awareness of self and other attending to same object)	Combination of dyadic and simple attending with mutual awareness (triadic relation)	Mutual shared attending as a means plus a mutually acknowledged purpose
Simple attention	X				
Dyadic attention		X			
Shared attention			X		
Mutual shared attention				X	
Joint attention					X