Neural Networks 23 (2010) 998–1003

Contents lists available at ScienceDirect

Neural Networks

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

2010 Special Issue

A minimal architecture for joint action

Cordula Vesper [a,*], Stephen Butterfill [b], Günther Knoblich [a], Natalie Sebanz [a]

[a] *Centre for Cognition, Donders Institute for Brain, Cognition, & Behaviour, Radboud University Nijmegen, The Netherlands*
[b] *Department of Philosophy, University of Warwick, United Kingdom*

ARTICLE INFO

Keywords:
Joint action
Coordination
Prediction
Monitoring
Task representation

ABSTRACT

What kinds of processes and representations make joint action possible? In this paper, we suggest a minimal architecture for joint action that focuses on representations, action monitoring and action prediction processes, as well as ways of simplifying coordination. The architecture spells out minimal requirements for an individual agent to engage in a joint action. We discuss existing evidence in support of the architecture as well as open questions that remain to be empirically addressed. In addition, we suggest possible interfaces between the minimal architecture and other approaches to joint action. The minimal architecture has implications for theorising about the emergence of joint action, for human–machine interaction, and for understanding how coordination can be facilitated by exploiting relations between multiple agents' actions and between actions and the environment.

1. Introduction

Joint action is distinct from individual action in a number of ways. Performing actions together often requires predicting what others are going to do next, adjusting one's behaviour to complement another's task, and achieving precise temporal coordination. For example, to play a piano duet a pianist must anticipate her partner's playing and adjust her own timing accordingly. What kinds of representations and processes make joint action possible?

The answers that have been given so far fall into two broad categories. Some researchers have emphasised the role of language (Clark, 1996), attributions of intention, belief and other propositional attitudes (Bratman, 1993, 1997, 2009; Tuomela, 2005), and commitments (Gilbert, 1992). Others have proposed that coordinated action fundamentally rests on direct perception-action links not requiring representations (Marsh, Johnston, Richardson, & Schmidt, 2009; Schmidt & Richardson, 2008). In our view, both approaches are illuminating but incomplete. Approaches that focus on language and propositional attitude ascription are well suited to cases in which long-term planning is involved or in which the agents engaged in joint action are not aware of the details of each other's actions, for example because they are separated in time or space. What these approaches do not explain is how precise temporal coordination, short-term adaptations to others' behaviour and short-term predictions about what another will do in the next (milli-) seconds are achieved. Approaches that focus on direct perception-action links within a dynamical systems framework are able to account for synchronisation of continuous movement. However, other aspects of joint action such as planning and prediction are not easily explained by this approach.

The aim of the present paper is to fill this gap by proposing an architecture that addresses the cognitive processes enabling people to perform actions together. It covers planning for immediate actions, action monitoring and action prediction, and ways of simplifying coordination. Unlike the dynamical systems framework that considers interpersonal coordination as a special case of more general coordination principles, the proposed framework assumes the existence of dedicated mechanisms for joint action. Unlike approaches focusing on language and shared intentionality that are mainly concerned with thinking and communicating about acting together, the framework is geared towards explaining how people actually perform actions together.

2. A minimal architecture for joint action

What kinds of processes and representations make performing joint actions possible? As a step towards answering this question, we propose a minimal architecture that could support joint action. This architecture consists of representations, processes and what we call coordination smoothers; together, these support basic forms of joint action. The minimal architecture does not attempt to define what joint action is and can be made compatible with various definitions. Instead of defining joint action, the minimal architecture aims at specifying building blocks that make performing joint actions possible.

---

* Corresponding address: Centre for Cognition, Donders Institute for Brain, Cognition, & Behaviour, Radboud University, PO Box 9104, 6500 HE Nijmegen, The Netherlands. Tel.: +31 24 36 12562.
E-mail address: c.vesper@donders.ru.nl (C. Vesper).

0893-6080/$ – see front matter © 2010 Published by Elsevier Ltd
doi:10.1016/j.neunet.2010.06.002

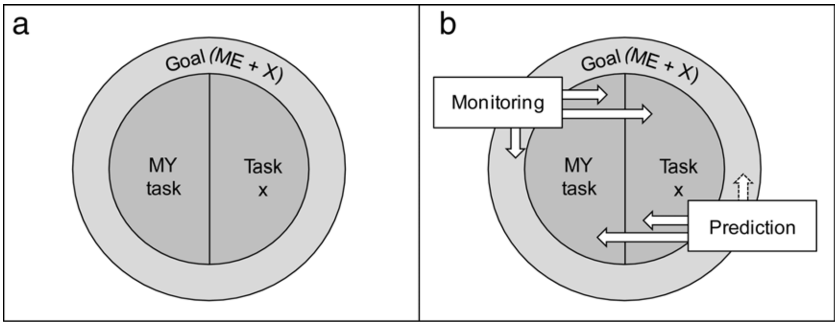

Fig. 1. (a) Representations supporting joint action can either refer just to one's own task and a goal that will not be achieved alone, or also include task x. Note that the goal need not involve more than the tasks. (b) Monitoring and prediction processes can act on representations of the goal, one's own task, and task x.

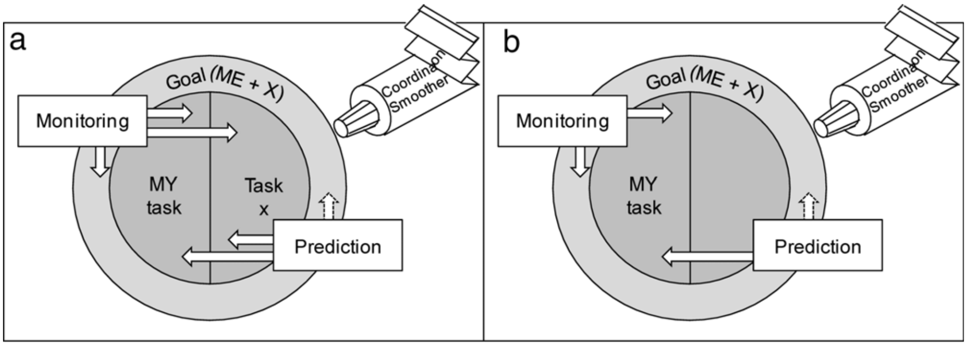

Fig. 2. (a) Coordination smoothers simplify coordination. (b) They are especially useful when actions pertaining to task x cannot be monitored or predicted.

Representations. In both individual and joint action, people represent goals and the tasks that need to be performed to achieve these goals. For example, consider a person lifting a two-handled picnic basket with another. Her goal is to put the basket into a car boot. What she needs to do – her task – is just to raise one of the handles in such a way that the other can synchronise with her. We assume that she represents the goal (moving the picnic basket) and her task. In basic cases, however, we propose that it is not necessary for her either to represent the other's task or to conceive of the other as an intentional agent. How does this joint situation differ from a case of individual action where a person lifts a basket alone? From the agent's point of view, the only difference is that the agent realises that performing her task is not sufficient for achieving her goal. In other words, given how the agent represents her task, the goal can only be achieved with the support of X, either another agent or some other force (Fig. 1(a)). This is captured by the formula ME + X.

In the simplest cases, we suppose that the agent represents only her own task. In many cases, though, it is necessary for the agent to represent another task, corresponding to what X is expected to do. This other task (task x) may be represented either as something that she will not do or as something that X is expected to do.

Processes. Two different processes operate on the representations identified above: monitoring and prediction (Fig. 1(b)). Monitoring processes determine to what extent a particular task or goal is being achieved and whether actions are unfolding as specified (Botvinick, Braver, Barch, Carter, & Cohen, 2001). On the extent to which tasks and goals are being achieved, three things could be monitored: the agent's own task, task x, and the goal. Depending on the kind of joint action performed, these may not all be monitored. To illustrate, it is perfectly possible for an agent to lift a two-handled basket even though she only monitors her own task and progress towards the goal and despite the fact that the lifting depends on another agent's cooperation. However, it is likely that in many cases monitoring task x will improve the performance in joint action. In the basket-lifting example, monitoring to what extent the other agent is succeeding by means of visual or haptic cues may facilitate synchronised lifting.

The prediction process is concerned with how actions will unfold and is often essential for precise coordination. Predictions about the immediate future are achieved through motor simulation whereby internal models specify how actions will affect the environment and what sensory consequences they will have (Wolpert & Ghahramani, 2000). As we will show later, such motor simulation can generate predictions about our own and others' actions (Wolpert, Doya, & Kawato, 2003). An open question is whether there are comparable mechanisms for predicting the combined outcomes of multiple agents' actions (Keller, 2008; Sebanz & Knoblich, 2009).

Coordination smoothers. Where joint action requires precise coordination in time or space, there are often limits on how well X's actions can be predicted. One way to facilitate coordination is for an agent to modify her own behaviour in such a way as to make it easier for others to predict upcoming actions, for example by exaggerating her movements or by reducing the variability of her actions. These are examples of coordination smoothers (see Fig. 2). In general, a coordination smoother is something that reliably has the effect of simplifying coordination (Fig. 2(a)). Coordination smoothers include both modulations of one's own behaviour (as in the above examples) as well as uses of objects that afford particular task distributions. The less an individual knows about task x or the way task x is performed, the more useful it is to deploy such coordination smoothers — in these cases, coordination smoothers may help to compensate for the lack of information (Fig. 2(b)).

3. Evidence

3.1. Representation

According to our architecture for joint action, an individual has to minimally represent her own task and the goal. It is not necessary, although typically very useful, to represent the other's task as well. To illustrate, consider two people dancing a tango together. In that case, the leader and follower have to coordinate their steps with each other. In principle, the follower might represent the leader's task in a very specific way, including for instance the other's need to make a forward movement with her right foot. This would imply that the follower could in principle switch roles with the leader. But it is more likely that the follower does not represent the leader's task in any such detail, and plausible that she does not represent the leader's task at all. All she requires is a representation of her own task and the goal of moving along a certain trajectory without losing contact between herself and her partner.

Can agents who do not represent each other's tasks really act in ways that jointly bring about a goal? Evidence would be provided by studies where an agent has no access to information about a cooperation partner's task but is nevertheless able to bring about a goal as a common effect of her and her partner's efforts. This relates to the question of whether collaboration in non-human animals and human children involves representing others' tasks. Existing studies of collaboration in chimpanzees could be interpreted as not involving representations of task x. For instance, Melis, Hare, and Tomasello (2006) presented chimpanzees with a food shelf that they could only access if two chimpanzees pulled on either end of a rope at approximately the same time. Chimpanzees were able to do this, and even to select the best coordination partner. However, it is possible that a chimpanzee did not represent the other's task and instead used her partner as a "social tool" (Moll & Tomasello, 2007). In particular, she may have realised that she needs to inhibit pulling on the rope except when it was tensing, which occurred when the other was starting to pull.

In a similar vein, it could be that the earliest joint actions infants perform together with adults do not require the infant to represent the other's task. For instance, an infant may have the goal to fetch a favourite toy from a high shelf and realise that she cannot get this done alone. The way to get the toy might involve the adult lifting the infant up so that she can reach for the toy. In this case, the infant does not need to represent what the adult is doing in order to achieve the goal together. However, there is evidence to suggest that, from an early age on, children are sensitive to others' parts in a joint action (Moll & Tomasello, 2007). They protest when an agent does not act in accordance with pre-specified rules (Rakoczy, Warneken, & Tomasello, 2008), and seem to have an understanding that acting together implies commitment (Carpenter, 2009; Gräfenhain, Behne, Carpenter, & Tomasello, 2009).

Although an agent involved in joint action might represent only her own task and the joint goal, as in the minimal case, it is generally useful to represent the other's task. This enables one to predict what the other will do next. In fact, it has been shown that adults are prone to represent what another is doing even when they are acting individually (in which case representing another's task can be detrimental to one's own performance). In studies on co-representation (Atmaca, Sebanz, Prinz, & Knoblich, 2008; Sebanz, Knoblich, & Prinz, 2003, 2005; Tsai, Kuo, Jing, Hung, & Tzeng, 2006), two people perform separate tasks alongside each other. A sequence of stimuli appears on a computer screen. One participant is instructed to respond to one type of stimulus, while her co-actor has to respond to another type of stimulus. Some of the stimuli are such that if the participant were performing both tasks alone, or if she were to represent the co-actor's task, then her performance would be disrupted. The results showed that the participant's performance in the two-person setting was disrupted in just the way it would be if she were performing both tasks alone. This suggests that people represent the other person's task, even when acting on a goal that the co-actor cannot contribute to. Further support for this view comes from electrophysiological evidence showing that people mentally perform the co-actor's task (Sebanz, Knoblich, Prinz, & Wascher, 2006; Tsai, Kuo, Hung, & Tzeng, 2008).

We propose that the co-representation identified in these tasks is a key ingredient that makes joint action possible in many cases. That is neither to say that co-representation is necessary in every case of joint action, nor that it only occurs in joint action settings. Indeed, the studies on co-representation indicate that representing others' tasks is pervasive, occurring outside of joint action.

So far we have been concerned with the representation of the task that the other has to perform (task x). In some cases agents may also represent certain properties of their partners (X). In one study (Richardson, Marsh, & Baron, 2007), pairs of participants had to pick up and move wooden planks of increasing or decreasing length. They could decide to pick them up individually or with a partner. The authors found that the point at which individuals shifted from using one hand to two hands, as measured by the ratio of plank length to hand span, was the same as the point at which participants switched from individual to joint lifting, as measured by the ratio of plank length to combined arm span. The present architecture would explain this finding as demonstrating that an individual represents some properties of her partner's body and the combined action-possibilities these entail.

In summary, agents may represent neither the other task to be performed (task x) nor anything about the performer of this task (X). However, in many cases, they will represent task x. This task could be represented in an agent-neutral way, without representing X, or be tied to a representation of (features of) X.

3.2. *Monitoring*

A monitoring process determines to what extent a particular task or goal is being achieved or whether actions are unfolding as specified (Botvinick et al., 2001). In some cases, joint action will only require monitoring the combined outcome of the agents' actions, but in other cases some of the agents involved in a joint action will also monitor each other's actions or the individual outcomes of those actions (see arrows in Fig. 1(b)). Consider two people playing a piano duet. If both are experts, they will need to monitor the music as the combined outcome of their effort (Keller, 2008), and perhaps also their own actions. But they will not necessarily monitor each other's actions or the individual outcomes of these actions. However, if in playing the duet one is teaching the other, the teacher is likely to monitor her pupil's actions in addition to monitoring the combined musical outcome.

In experimental studies, monitoring is often operationally defined as error detection. Behavioural markers of error detection include slowing down after making an error (Rabbitt, 1966). Electrophysiological activation, such as a negative brain potential building up as errors are made, marks error detection as well (Falkenstein, Hoormann, Christ, & Hohnsbein, 2000). These markers are also found when people observe errors made by others (Schuch & Tipper, 2007; van Schie, Mars, Coles, & Bekkering, 2004), suggesting that the same monitoring mechanisms may be used for one's own and others' actions. Further evidence comes from brain imaging studies showing that similar brain regions are active when processing one's own and others' errors (de Bruijn, de Lange, von Cramon, & Ullsperger, 2009; Malfait et al., 2009; Newman-Norlund, Ganesh, van Schie, Bruijn, & Bekkering, 2009).

Given that most of these studies have investigated monitoring in the context of action observation rather than joint action, it is an open question to what extent the combined outcome of one's own and others' actions is monitored. Direct evidence for such monitoring would be provided if agents involved in a joint action were more sensitive to errors in their combined performance than to errors in their individual performance. This predicts that if two agents' individual errors cancelled each other out, they should not become aware of having made an error at all. As far as we know no such direct evidence is available.

Although it is not generally necessary for joint action, monitoring the other's task can greatly improve performance. In a study investigating musical coordination (Goebl & Palmer, 2009), two pianists played a piano duet together under three conditions: they received normal auditory feedback; they only heard the partner's playing; or they only heard their own playing. Coordination performance, as measured by the asynchrony between tones, was best with normal auditory feedback and worst in the condition in which participants only heard their own tones but not the other's. This indicates that being able to monitor the other's actions facilitates coordination.

3.3. *Prediction*

Prediction is concerned with the unfolding of actions in the immediate future. In contrast to predictions about more long-term events (such as who will buy sheet music for a piano duet), the process targeted here generates expectations about the outcome of actions online (such as when will one's duet partner play the next tone). This kind of prediction relies on motor simulation whereby internal models specify the immediate consequences of one's actions.

Internal predictive models were first described for individual action, as they are useful for individual motor control (Wolpert & Ghahramani, 2000). Neurophysiological studies suggest that these models are mainly located in the cerebellum (Wolpert, Miall, & Kawato, 1998) and are continuously updated by comparing the actual and predicted consequences of one's actions. It has since been suggested that internal models can also make predictions about others' actions in the immediate future (Wolpert et al., 2003). Evidence for this claim comes from studies demonstrating that people are able to anticipate the future course of others' actions (Graf et al., 2007). Running simulations of others' actions as they unfold may be especially useful in the context of joint action as it is thought to bias perception and to help with anticipating the consequences and, most importantly, the timing of others' actions (Wilson & Knoblich, 2005).

The view that internal predictive models are involved in the ability to anticipate others' actions is further supported by studies showing that short-term predictions of others' actions rely on one's own motor experience (e.g. Aglioti, Cesari, Romani, & Urgesi, 2008; Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005). For instance, infants able to crawl showed more evidence for motor simulation when observing other infants crawling, compared to seeing them walking (van Elk, van Schie, Hunnius, Vesper, & Bekkering, 2008). The fact that prediction of this kind involves motor simulation also explains why people are able to recognise their own earlier actions based on temporal cues (Flach, Knoblich, & Prinz, 2004) and are most accurate at generating predictions about their own actions (Knoblich & Flach, 2001).

What does this imply for joint action? Generally, joint action coordination should be better the more accurately interaction partners can predict the timing of each other's actions. This implies that performance would be best if one could interact with oneself because, given one's motor experience, one's own actions should be easiest to predict. Keller, Knoblich, and Repp (2007) tested this prediction by comparing how well professional pianists synchronised their performance in a duet when playing with recordings of either another pianist or an earlier recording of themselves. Indeed, the results showed better coordination performance (measured by the synchronisation error for single tones) when pianists played with their own recordings. This finding indicates that more accurate predictions about future events produced by oneself and others may lead to better interpersonal coordination.

While there is evidence to show that an agent can make specific predictions about the immediate actions of her partner (predicting the unfolding of task x online, see Fig. 1(b)), it remains to be determined whether predictions are also made regarding the consequences of the combined actions of the agent and her partner (Keller, 2008; Sebanz & Knoblich, 2009). For instance, will a pianist playing a duet make predictions about the other's tones, or will the predictions be about the music, reflecting the combined outcome of her own and the other's playing? Some evidence for the prediction of combined outcomes is provided by a study (Knoblich & Jordan, 2003) in which participants jointly controlled a cursor with the aim of tracking a moving target on a computer screen. The results showed that participants initially found this task quite demanding and were worse than single participants performing the task alone. However, with practice, joint task performance reached the level of individual performance. This suggests that participants learned to predict how their own and the other's actions would jointly affect the movement of the tracker.

3.4. *Coordination smoothers*

A coordination smoother is either a modulation of one's own behaviour which reliably has the effect of simplifying coordination, or the use of an object that affords a particular task distribution which reliably has such an effect. One type of coordination smoother involves making one's own behaviour more predictable. For instance, speeding up has the effect of making movements less variable. A recent study (Vesper, van der Wel, Knoblich, & Sebanz, submitted for publication) where participants were instructed to synchronise responses found that partners whose actions were less variable achieved better coordination. In this study, participants decreased their variability by speeding up their own responses.

A second type of coordination smoother involves ways of delimiting and structuring one's own task such that the need for coordination is reduced. In a joint search activity, delimiting might mean confining one's searching to the most obvious region of space for one to search in. This is indicated by a study in which pairs of participants jointly performed a visual search task (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008). Performance was best when participants received feedback about the other's search. This might have allowed them to divide up the search space more effectively. Interestingly, verbal communication did not contribute to improving joint performance. In a joint building task, delimiting one's own task might mean moving in a way least likely to enter another's action space (Vesper, Soutschek, & Schubö, 2009). Imposing structure on a task might involve trying to relate one's actions to the other's actions in time. Turn-taking is a paradigm example: someone picking apples from a tree with a friend might introduce a delay between her attempts to pick apples so that the other can have a turn, which would avoid interference resulting from two people moving the branches of the tree simultaneously.

A further type of coordination smoother involves providing coordination signals (Clark, 1996). Some coordination signals are conventional, such as musical scores and traffic signs. More relevant for the aspects of joint action at issue here are nonconventional signals. Agents are able to selectively make certain movements salient with the effect that information about their actions is more readily available to others; in this case, movements serve as both components of actions and coordination signals. One such example is provided by the earlier mentioned study by Goebl and Palmer (2009) in which pianists played a duet under varying feedback conditions. Analyses of participants' finger movements showed that they raised their fingers higher and synchronised their head movements more when the auditory feedback was reduced. This is in line with our prediction that coordination smoothers become more important as less is known about task x (see Fig. 2(b)).

Synchronisation could also serve as a coordination smoother. For example, Wilson and Wilson (2005) have proposed that a common speech rhythm could help conversation partners to precisely predict when an utterance will end. This enables seamless turn-taking of speaker and listener. The same idea can be applied to nonverbal coordination. One could speculate that synchronisation of movements makes interaction partners more similar and thus more predictable to one another (see also Keller et al., 2007) insofar as prediction of others' actions relies on motor simulation. As an example, when two people in rocking chairs fall into synchrony (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007) this might make it easier for one of them to predict when exactly the other will pass her a mug of hot chocolate.

In addition to modulations of one's own behaviour, objects affording a particular task distribution can also serve as coordination smoothers. While some objects afford a particular use by single individuals (Gibson, 1977), other objects afford joint use by virtue of their size, form, weight, and so on. Some of these may be such that they afford a particular task distribution, or a particular way of handling them together. For instance, a two-handled picnic basket affords that the left agent grasps the left handle with her right hand, and that the right agent grasps the right handle with her left hand. As far as we know, it remains to be investigated empirically whether object affordances act as coordination smoothers and which object features are crucial in constraining joint action.

4. Interfaces

The aim of the present paper was to spell out a minimal architecture for joint action that focuses on planning for immediate actions, action monitoring and action prediction, as well as ways of simplifying coordination. How does our architecture relate to approaches that focus on direct perception-action links within a dynamical systems framework, and how is it linked to approaches that focus on language and propositional attitudes?

On the one hand, it has been proposed that coordinated action rests on direct perception-action links not requiring representations (Marsh et al., 2009; Schmidt & Richardson, 2008). This view is embedded in dynamical systems theory, which has been able to explain a wide range of synchronised collective behaviours, from the formation of flying birds (Couzin, Krause, Franks, & Levin, 2005) to the synchronised swaying of people in conversation (Shockley, Baker, Richardson, & Fowler, 2007). These phenomena are not the target of our minimal architecture. However, they are relevant to understanding certain aspects of joint action. Can there be crosstalk between a dynamical systems view and our architecture? The minimal architecture assumes that synchronisation of movements serves as a coordination smoother that makes oneself or the other more predictable. While it is beyond our architecture to characterise how movements become synchronised, dynamical systems theory offers an account to explain how this is possible. In this way, findings from dynamical systems theory complement our architecture. Conversely, our minimalist architecture offers a ladder from synchronised movement to the sorts of joint action that have proven difficult to capture within a radical dynamical systems framework — joint actions involving discrete action, complementary action, and prediction.

On the other hand, in explaining how joint action is possible, some researchers have emphasised the role of language (Clark, 1996), attributions of intention, belief and other propositional attitudes (Bratman, 1993, 1997; Tollefsen, 2005; Tuomela, 2005), as well as commitments (Gilbert, 1992). How does our minimal architecture relate to these approaches? A limit of our architecture is that it does not generate precise predictions about how joint actions involving long-term planning will unfold. These are the cases that language-and-intention approaches are best suited to explain. Rather than extend the minimal architecture to encompass such cases, it seems more useful to specify connections between the mechanisms we have discussed and those highlighted by approaches emphasising language and intention. Identifying such connections would also help to overcome a limit of the latter approaches. In many cases of joint action, it is insufficient to agree on a course of action: there is also the issue of performing the planned actions together. These approaches do not make specific predictions regarding actual performance. Regardless of how much people talk and how deeply interlocked their intentions, performing a piano duet together will also require exquisitely coordinated timing. Such coordination is not readily explained by shared intention or linguistic communication. Our minimal architecture fills this gap: given that linguistic communication and shared intentions can guide monitoring and prediction processes and influence coordination smoothing, the architecture identifies mechanisms which make possible the precise coordination that is so often essential for implementing shared plans.

References

Aglioti, S. M., Cesari, P., Romani, M, & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*(9), 1109–1116.

Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action co-representation: the joint SNARC effect. *Social Neuroscience*, *3*(3–4), 410–420.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

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

Bratman, M. E. (1997). I Intend that we J. In R. Tuomela, & G. Holmstrom-Hintikka (Eds.), *Contemporary action theory, volume 2: social action* (pp. 49–63). Dordrecht: Kluwer.

Bratman, M. E. (2009). Modest sociality and the distinctiveness of intention. *Philosophical Studies*, *114*(1), 149–165.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: the costs and benefits of shared gaze during collaborative search. *Cognition*, *106*, 1465–1477.

Calvo-Merino, B., Glaser, D., Grèzes, J., Passingham, R., & Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cerebral Cortex*, *15*, 1243–1249.

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.

Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision making in animal groups on the move. *Nature*, *433*, 513–516.

de Bruijn, E. R. A., de Lange, F. P., von Cramon, D. Y., & Ullsperger, M. (2009). When errors are rewarding. *Journal of Neuroscience*, *29*(39), 12183–12186.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology*, *51*, 87–107.

Flach, R., Knoblich, G., & Prinz, W. (2004). Recognizing one's own clapping: the role of temporal cues. *Psychological Research*, *69*, 147–156.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw, & J. Bransford (Eds.), *Perceiving, acting, and knowing: toward an ecological psychology* (pp. 67–82). Hillsdale, NJ: Erlbaum.

Gilbert, M. (1992). *On social facts*. Princeton, NJ: Princeton University Press.

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, *26*(5), 427–438.

Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., & Prinz, W. (2007). Predicting point-light actions in real-time. *Neuroimage*, *36*, T22–T32.

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. *Developmental Psychology*, *45*(5), 1430–1443.

Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: a cognitive and social perspective on the study of interactions* (pp. 205–221). Amsterdam: IOS Press.

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Consciousness and Cognition*, *16*, 102–111.

Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: interactions of perception and action. *Psychological Science*, *12*(6), 467–472.

Knoblich, G., & Jordan, J. S. (2003). Action coordination in groups and individuals: learning anticipatory control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 1006–1016.

Malfait, N., Valyear, K. F., Culham, J. C., Anton, J.-L., Brown, L. E., & Gribble, P. L. (2009). FMRI activation during observation of others' reach errors. *Journal of Cognitive Neuroscience*, *22*(7), 1493–1503.

Marsh, K. L., Johnston, L., Richardson, M. J., & Schmidt, R. C. (2009). Toward a radically embodied, embedded social psychology. *European Journal of Social Psychology*, *39*, 1217–1225.

Melis, A. P., Hare, B., & Tomasello, M. (2006). Chimpanzees recruit the best collaborators. *Science*, *311*, 1297–1300.

Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society of London B*, *362*, 639–648.

Newman-Norlund, R. D., Ganesh, S., van Schie, H. T., Bruijn, E. R. A. D., & Bekkering, H. (2009). Self-identification and empathy modulate error-related brain activity during the observation of penalty shots between friend and foe. *Scan*, *4*, 10–22.

Rabbitt, P. M. A. (1966). Errors and error correction in choice–response tasks. *Journal of Experimental Psychology*, *71*(2), 264–272.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. *Developmental Psychology*, *44*(3), 875–881.

Richardson, M. J., Marsh, K. L., & Baron, R. M. (2007). Judging and actualizing intrapersonal and interpersonal affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 845–859.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*, 867–891.

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs, & V. K. Jirsa (Eds.), *Coordination: neural, behavioral and social dynamics* (pp. 281–308). Berlin, Heidelberg: Springer.

Schuch, S., & Tipper, S. P. (2007). On observing another person's actions: influences of observed inhibition and errors. *Perception and Psychophysics*, *69*(5), 828–837.

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: what, when, and where. *Topics in Cognitive Science*, *1*, 353–367.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, *88*, B11–B21.

Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: corepresenting stimulus–response mappings. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1234–1246.

Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: an ERP study of action planning and control in coacting individuals. *Journal of Cognitive Neuroscience*, *18*(5), 859–870.

Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 201–208.

Tollefsen, D. (2005). Let's pretend! children and joint action. *Philosophy of the Social Sciences*, *35*(1), 75–97.

Tsai, C.-C., Kuo, W.-J., Hung, D. L., & Tzeng, O. J. L. (2008). Action co-representation is tuned to other humans. *Journal of Cognitive Neuroscience*, *20*(11), 2015–2024.

Tsai, C.-C., Kuo, W.-J., Jing, J.-T., Hung, D. L., & Tzeng, O. J.-L. (2006). A common coding framework in self-other interaction: evidence from joint action task. *Experimental Brain Research*, *175*, 353–362.

Tuomela, R. (2005). WE-intentions revisited. *Philosophical Studies*, *125*, 327–369.

van Elk, M., van Schie, H., Hunnius, S., Vesper, C., & Bekkering, H. (2008). You'll never crawl alone: neurophysiological evidence for experience-dependent motor resonance in infancy. *Neuroimage*, *43*, 808–814.

van Schie, H. T., Mars, R. B., Coles, M. G. H., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, *7*, 549–554.

Vesper, C., van der Wel, R.P.R.D., Knoblich, G., & Sebanz, N. Better be fast: speeding supports coordination of intermittent joint actions (submitted for publication).

Vesper, C., Soutschek, A., & Schubö, A. (2009). Motion coordination affects movement parameters in a joint pick-and-place task. *Quartely Journal of Experimental Psychology*, *62*(12), 2418–2432.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*(3), 460–473.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychological Bulletin and Review*, *12*(6), 957–968.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and interaction. *Philosophical Transactions of the Royal Society of London B*, *358*, 593–602.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, *3*, 1212–1217.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*(9), 338–347.