Prediction in Joint Action: What, When, and Where

Natalie Sebanz, Guenther Knoblich

Centre for Cognition, Donders Institute for Brain, Cognition, & Behaviour, Radboud University Nijmegen

Abstract

Drawing on recent findings in the cognitive and neurosciences, this article discusses how people manage to predict each other's actions, which is fundamental for joint action. We explore how a common coding of perceived and performed actions may allow actors to predict the what, when, and where of others' actions. The ''what'' aspect refers to predictions about the kind of action the other will perform and to the intention that drives the action. The ''when'' aspect is critical for all joint actions requiring close temporal coordination. The ''where'' aspect is important for the online coordination of actions because actors need to effectively distribute a common space. We argue that although common coding of perceived and performed actions alone is not sufficient to enable one to engage in joint action, it provides a representational platform for integrating the actions of self and other. The final part of the paper considers links between lower-level processes like action simulation and higher-level processes like verbal communication and mental state attribution that have previously been at the focus of joint action research.

Keywords: Joint action; Prediction; Perception–action links; Action simulation; Mirror system; Joint attention; Mental state attribution

1. Introduction

It seems that much of our ubiquitous ''social nature'' as humans plays out in virtual space these days: We e-mail friends rather than taking a walk with them around the village, we use Internet sites to compare the prices for goods rather than haggling over the price of an object at the market, and we develop and employ robots and other machines to do the work that in earlier times may have required cooperation of several individuals, such as plowing a field or building a car. This may lead to the impression that what matters about social

Correspondence should be sent to Natalie Sebanz, Centre for Cognition, Donders Institute for Brain, Cognition, & Behaviour, Radboud University, PO Box 9104, 6500 HE Nijmegen, The Netherlands. E-mail: n.sebanz@donders.ru.nl

interaction primarily has to do with thinking about others, attributing mental states to them, making moral and economic judgments, and engaging in symbolic communication. However, neither as individuals nor as a species would we exist were it not for our ability to engage in joint action with others (Tomasello, Carpenter, Call, Behne, & Moll, 2005; Vygotsky, 1987). Being able to coordinate one's actions with others in time and space means that the scope of potential products of action dramatically increases (Clark, 1996).

While interest in human forms of joint action has shaped research into anthropology (e.g., Henrich & McElreath, 2003), pragmatics (e.g., Fetzer, 2007), and sociolinguistics (e.g., Gumperz, 1982), the majority of cognitive scientists has tended to picture humans as being highly autonomous and individualistic, and to characterize social processes as special cases of individual thinking. A few notable exceptions include work on distributed cognition (e.g., Hutchins, 1995). Currently, however, there is a new trend in the cognitive sciences to study joint action with the perspective that basic forms of interaction provide important social constraints for the architecture of individual cognitive processing (Sebanz, Bekkering, & Knoblich, 2006a). This is partly due to an increased interest in the cognitive and neural mechanisms underlying commonalities in action production and action perception (Prinz, 1990, 1997) and the processes that allow us to coordinate our actions with those of others (Knoblich & Jordan, 2003). In this article, we consider how recent research in the cognitive and neurosciences can inform our understanding of joint action. We focus on the question of how individuals manage to predict each other's actions, which is critical for joint action. If we merely reacted to what we saw others doing, we could never achieve the smooth and fast coordination needed to pass a ball, to jointly lift a tray full of plates and glasses, or to drive in heavy traffic.

Three critical aspects about others' actions that predictions can deliver are what, when, and where others will act. The ''what'' aspect refers to predictions about the kind of action the other will perform and to the intention that drives the action. If our goal is to jointly change the bed linens, we want to know whether our partner will start with a pillow or get the blanket changed first. The ''when'' aspect is critical for all joint actions requiring close temporal coordination. For example, when trying to spread the newly covered blanket evenly over the bed, the two individuals grasping the blanket at opposite ends must try to lift it and lower it at the same time. Otherwise, it may slip from one actor's hands while the other is lifting it. Temporal coordination is even more critical in activities such as rowing a boat together or making music (Keller, 2008). Finally, the ''where'' aspect is also important for the online coordination of actions because actors need to effectively distribute a common space to avoid collisions and optimize movement paths (Erlhagen & Bicho, 2006). To revisit our example, it is easiest to jointly lift a blanket when the two co-actors grasp it at opposite ends.

Previous research has focused on verbal communication as a powerful device for agreeing on how a joint action will be performed (Clark, 1996). However, verbal exchange alone is sometimes not sufficient to achieve the fine-grained attunement required for joint actions. For example, if a soccer team agrees prior to a game on how they will try to score a goal, this is no guarantee that the striker will get to the ball in the time window of a few hundred milliseconds during which the planned opportunity arises.

During the last decade, research in cognitive psychology and cognitive neuroscience has accumulated evidence for links between people's bodies in action that are more direct and, by implication, faster than verbal/symbolic communication. In particular, it has been found that we rely on our own motor system when perceiving and predicting others' actions (for a review, see Rizzolatti & Craighero, 2004). We will discuss the extent to which these links could solve coordination problems in joint action that have to do with predicting the what, where, and when of others' actions and the common results of joint action. In the last section of this paper, we will consider how such basic links between perception and action could work in concert with coordination mechanisms in the realm of language and mental state attribution.

2. Interpersonal links through action simulation

The groundwork for the idea that perceiving others' actions does not leave one's own motor system ''cold'' was laid by ideomotor theory (James, 1890; Lotze, 1852). This theory postulated that imagining an action (such as getting out of bed) would create a tendency to perform this action (and thus allow one to get out of bed). Many years later, cognitive psychologists realized that there is an interesting way to extend this functional principle to social interaction. This was achieved by replacing the act of imagining one's own future action with the act of perceiving someone else's action (Greenwald, 1970). So the new claim was that not only when we imagine an action, but whenever we observe someone performing an action, this will create a tendency to perform this action (Prinz, 1997).

Before we consider the empirical evidence for this claim and go on to discuss its implications for joint action, let us consider how a cognitive system would need to be structured in order to allow for such a link. Clearly, there must be a way in which observing another's action directly speaks to one's own motor system. It makes sense to assume, then, that one's own and others' actions should be coded in a common representational domain (Prinz, 1990, 1997). It should almost be as if planning to perform an action and seeing it performed were one and the same thing.

Evidence for this notion of ''common coding'' comes from behavioral studies as well as from neuroscience. To summarize the behavioral findings very briefly, it has been shown that (a) perceiving an action affects one's own concurrent performance of a related action (for a recent study, see Stanley, Gowen, & Miall, 2007); (b) even when people do not intend to perform a particular action, seeing others engage in it leads to a tendency to perform the same action (Lakin & Chartrand, 2003); and (c) perceptual judgments about others' actions are affected by the current state of one's own motor system (for a review, see Schütz-Bosbach & Prinz, 2007).

Through the use of single-cell studies in monkeys and various neuroscience techniques in humans, it has been possible to show that (a) single cells (mirror neurons) in monkey premotor and parietal cortex fire when the monkey observes an action like grasping a peanut and when the monkey performs the same action himself or herself; (b) areas in premotor and parietal cortex in humans show similar activation when individuals act and when they

observe others' actions; (c) the amount of observation-induced activation depends on motor expertise, such that a dancer would show more activation in premotor cortex when seeing familiar dance movements compared to seeing unfamiliar ones; and (d) specific patterns in brain activity associated with action planning and performance are also found during observation (for references, see Grafton & Hamilton, 2007).

On the one hand, the results from these neuroscience studies have provided important additional evidence for functional principles that are difficult to test. On the other hand, it has to be said that these findings have sparked such intense interest and debate in and of themselves that we have almost lost sight of the functional principles postulated before and independently of the discovery of mirror neurons. We believe that it is important for psychologists to refocus on the functional assumption of a common coding for perception and action instead of getting caught in neophrenological ''brain'' discourses, at least, if the goal is to work out the social implications of the preverbal interpersonal link that common coding (mirroring) provides.

3. What: Inferring and predicting each other's goals

To engage in joint action, we must be able to understand what others are doing and to be able to predict what they will do next. Can the perception-action links sketched above provide the functionality for inferring and predicting action goals? Among the different functions ascribed to mirror neurons and to perception-action links more generally, the dominating claim is that perception-action links allow us to infer the goals of others' actions (Rizzolatti & Craighero, 2004). This claim rests on the finding that mirror neurons code the outcome of actions rather than the means by which actions are accomplished. For example, a cluster of mirror neurons may fire whenever the monkey sees someone grasping a piece of food, regardless of the exact movement trajectory. Likewise, mirror neurons fire when the monkeys hears a sound associated with a goal-directed action, such as breaking a peanut (Kohler et al., 2002). In line with this view, there is much evidence suggesting that humans code others' actions in terms of their goals, for example, when imitating actions (Bekkering, Wohlschläger, & Gattis, 2000).

A matter of debate is whether such goal inference is predictive or postdictive. Do we observe others' actions and then infer what their goal must have been, or do we predict the outcomes of their actions as they unfold? This question is of some importance because the first interpretation brings ''mirroring'' closer to the realm of understanding mental states, whereas the latter view stresses its general function in predicting a range of different action aspects, which has been suggested to improve the perceptual processing of ongoing actions (Wilson & Knoblich, 2005). In any case, it is important to stress that a process of ''mirroring'' through close perception-action links does neither require nor provide an understanding that other individuals have mental states, let alone an understanding that these mental states can differ from one's own (Knoblich & Jordan, 2002; Pacherie & Dokic, 2006).

So what can common coding do for joint action? Making sense of others' actions is of course essential for joining in when others are already engaged in an activity, as well as for

being able to adjust one's actions depending on what one's interaction partner is doing. It seems unlikely, however, that individuals whose social abilities are restricted to shared representations for perception and action can fully exploit these advantages in joint action (Knoblich & Sebanz, 2008). The reason for this is that intentional joint action—the deliberate attempt to coordinate one's actions with others' to bring about a change in the environment—seems to require a representation of the joint task that involves one's own and the other's part. This, in turn, may require that one is able to conceive of others as (intentional) agents like oneself (see also Call, this issue, pp. 368–379; Carpenter, this issue, pp. 380–391).

Accordingly, we will now consider what common coding can contribute to joint action once the ability to treat self and other as independent intentional agents is already in place. We would like to suggest two possible functions: First, if action simulation is conceived as a process that can become decoupled from the actual observation of the other's actions, knowing the other's part in a task through symbolic cues, verbal communication, or conventions can be sufficient to generate action-related predictions. In support of this assumption, it has been found that when observers are told under which conditions a certain action is to be expected, motor activation is observed even when another's actions are replaced by a symbolic cue announcing his or her actions (Kilner, Vargas, Duval, Blakemore, & Sirigu, 2004; Ramnani & Miall, 2004).

In a series of behavioral experiments, we showed that knowing a co-actor's task affects one's own action planning and performance even when there is no need to take the other's part into account at all (Atmaca, Sebanz, Prinz, & Knoblich, 2008; Sebanz, Knoblich, & Prinz, 2003, 2005). The logic of these experiments was to compare whether a particular pattern that is typically observed when a single person performs a task can also be observed when the task is distributed across two individuals. It is known that when a single person performs a task that involves repeatedly choosing between two actions (such as a left and a right key press) in response to stimuli, one can easily induce a conflict in action selection by adding an irrelevant feature to the stimuli that, despite being task irrelevant, tempts one into performing a left or right action. For example, when your task is to respond to pictures of a red ring with a left key press, and to pictures of a green ring with a right key press, adding a finger that points left or right makes it difficult to respond when the finger does not point in the same direction as the response required by the ring.

Can such a conflict also be observed when two co-actors perform the task together, so that each is responsible for only half of the task? If they successfully ignored each other, no conflict should occur, because each actor is in charge of one action alternative only. However, if each co-actor represented the other's action alternative in a similar way as his or her own, there should be a similar conflict in action selection as when a single person performs the whole task. This was indeed found. Participants were faster to respond to their ring color when the finger pointed toward them, than when the finger pointed toward the other participant, thus activating a representation of the other's action. This effect was not observed in a control condition where participants performed their part of the task alone.

Moreover, when we looked at brain activation on so-called ''nogo trials,'' where participants should not act because it is not their turn, we found that more efforts to inhibit one's

tendency to act were made during the joint performance compared to the individual performance of one's part (Sebanz, Knoblich, Prinz, & Wascher, 2006b). This suggests that participants anticipated the other's action, which increased their own tendency to act. So these findings demonstrate first, that people have a strong tendency to form shared task representations, taking into account what those around them need to be doing, and second, that there is a level at which one's own and the co-actor's actions are represented in a functionally equivalent way, as proposed by common coding.

More generally, we can say that being able to run simulations of others' actions in one's own motor system that are guided by one's representation of the other's task allows one to anticipate the other's action and to imagine the other's action when it cannot be observed. In line with this assumption, we found that the joint action selection effects describe above occurred even when people could not observe each other's actions at all. Furthermore, there is evidence that the other's actions are only simulated when the other is perceived as acting intentionally. No attention is paid to the other's ask when his or her actions are controlled by a machine (S. Atmaca, N. Sebanz, and G. Knoblich, unpublished data). This indicates that action simulation does not occur always, per default, but rather, is constrained and recruited in the service of higher-level task representations. These representations are likely to be structured along self and other (Roepstorff & Frith, 2004). In line with this idea, we found that a brain area in medial prefrontal cortex, associated with self-awareness, showed increased activation when participants performed their part in a joint task, compared to performing the same part alone (Sebanz, Rebbechi, Knoblich, Prinz, & Frith, 2007). Thus, at the task level, self and other are kept apart, but the task representation recruits an underlying common coding system where the actions of self and other are represented in the same way.

A second way in which common coding can contribute to joint action is by providing a representational level supporting the monitoring and detection of others' errors using one's own action repertoire (see Bekkering et al., this issue, pp. 340–352). Knowing the other's task allows one to compare a simulation of what the other should do with the other's ongoing actions. In the context of joint action this means that one can rapidly adjust to or compensate for others' errors, or even initiate actions to avoid them. Finally, if the other's actions deviate from one's predictions but yield better results, this could be taken as a clue to update one's representation of a task or subtask to increase the efficiency of one's own actions. Such a mechanism could also provide a principle for deciding when to imitate others.

4. When: Becoming entrained and getting ahead of each other

Bringing one's own and another's action plans together in the service of a joint task representation is only one aspect of joint action. Many joint actions pose an additional challenge: predicting the timing of others' actions to achieve temporal coordination. This is crucial for acting synchronously or in turns. Time places serious constraints on joint action, and the time windows for coordination are often very narrow (a few hundred milliseconds or less). When we perform an action that requires bimanual coordination, such as

juggling balls in the air, we use one and the same system to coordinate the timing of the left and right hand. In contrast, when we perform the same task together with another person, the question arises as to how we manage to take the timing of the other's actions into account.

Research guided by the framework of ecological psychology has highlighted the role of entrainment. Even when people are asked *not* to synchronize with another while performing rhythmic actions like swinging a pendulum or rocking in a rocking chair, they show a tendency to entrain, that is, to fall into the same rhythm (see Marsh et al., this issue, pp. 320–339). The mechanism proposed to underlie entrainment relies on the notion of coupled oscillators. In brief, it is thought that entrainment occurs as a consequence of direct, unmediated perception-action links between two or more systems that become coupled. In line with the credo of ecological psychology that any reference to mental states should be avoided, the same mechanism is assumed to explain how clocks hanging on the same wall synchronize (Huygens, 1665), how two limbs become coupled (Schmidt, Bienvenu, Fitzpatrick, & Amazeen, 1998), how people adjust their rocking frequencies when sitting in rocking chairs (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007), and how crowds come to clap in unison (Neda, Ravasz, Brechte, Vicsek, & Barabasi, 2000).

The role of entrainment in intentional joint action is relatively unexplored (but see Shockley, Richardson, & Dale, this issue, pp. 305–319). One established finding is that entrainment occurs regardless of whether people are instructed to synchronize. Furthermore, recent findings suggest that entrainment and liking are related (Ouillier, DeGuzman, Jamtzen, Lagarde, & Kelso, 2008). Entrainment may lead to better affiliation between individuals, which in turn could facilitate intentional action coordination. It may even be used as a diagnostic tool to identify those individuals with whom one ''gels.'' Although a potential link between entrainment and affect is very interesting, a specific theoretical framework that would predict such relationships is still to be developed. Such a framework could also help to relate findings on entrainment and liking and the growing literature on nonconscious mimicry that has demonstrated links between affiliation and the tendency to mimic others' actions and mannerisms (Lakin & Chartrand, 2003). So far, studies on nonconscious mimicry have investigated how adapting the same behavior affects liking, but they have not investigated the temporal synchronization of actions, whereas studies on entrainment have exclusively focused on performance of the same, rhythmic actions (for a recent exception, see Watanabe, 2008). Generally, it seems high time to cross boundaries between psychological subdisciplines to investigate how these two phenomena are related.

Unless the ecological approach to entrainment can be extended to account for joint actions involving discrete and object-directed actions rather than highly rhythmic movements, it is confined to providing a systematic way of dealing quantitatively with a particular interpersonal phenomenon. In particular, the principle of coupled oscillators cannot explain how individuals flexibly adjust the timing of their actions to one another to achieve a common outcome in real time, such as when dancing together. Can action simulation provide a potential solution for this problem?

So far, we have only discussed the possibility that by simulating others' actions we can predict the outcome of observed actions. However, common coding may also provide the

necessary platform for simulating the timing of others' actions. Note that common coding itself only postulates that when observing an action, a corresponding representation of this action will be activated in the observer. Thus, there is a match at the level of action goals. For simulation to entail timing, one needs to additionally postulate mechanisms that allow one to apply temporal predictions generated in one's own motor system to observed actions. How could this be achieved?

According to current theories of motor control, we generate precisely timed predictions about the sensory consequences of our actions whenever we plan to move (relying on so-called forward models, Davidson & Wolpert, 2003). If our actions unfold as planned, the predicted and actual sensory consequences of our actions match, leading to what is known as ''sensory cancellation''—a filtering out of expected effects. This can explain why we cannot tickle ourselves (Blakemore, Frith, & Wolpert, 1999). When we try to tickle our-selves, we can exactly predict when we will touch ourselves, leading to sensory cancella-tion. In contrast, when the touch occurs later than expected, it becomes more ticklish, because the timing of the sensory consequences is off.

A recent study by Sato (2008) provides an indication that we may be able to accurately apply temporal predictions generated in our own motor system to observed actions (Wilson & Knoblich, 2005). Participants perceived a tone as less intense when they had just pressed a button to produce it, compared to when the tone occurred unexpectedly. Interestingly, a similar effect of sensory cancellation was found when participants observed another person pressing a key to produce the tone. So tones also appeared less intense to them when they had a chance to observe another person pressing the key to produce the tone. In all studies so far, sensory cancellation relied on accurate temporal predictions. Therefore, it is tempting to assume that the same holds for sensory cancellation during action observation. However, this needs to be further explored.

Another indication that real-time predictions contribute to action observation comes from studies investigating how people manage to recognize their own earlier actions based on purely temporal cues. One surprising result was that when people are asked to recognize recordings of clapping it does not make a difference whether they hear the original record-ing or sine tones that reflect the temporal pattern of the clapping (Flach, Knoblich, & Prinz, 2004). Furthermore, temporal aspects seem to be crucial to allow people to recognize earlier recorded, dynamic traces of their own handwriting (Knoblich & Prinz, 2001). Finally, expert pianists are very well able to recognize their own earlier performances of unfamiliar pieces based on expressive timing (Repp & Knoblich, 2004), that is, stable timing-deviations (in the order of tens of milliseconds) from the notated score. This suggests that they covertly play along with the performance and notice temporal deviations from how they would perform the piece.

Assuming that we rely on predictive mechanisms in our own motor system when simulat-ing others' actions, how does this affect joint action? In principle, there are two ways in which such real-time simulations could play out. On the one hand, it may be possible to run multiple parallel simulations (Keller, Knoblich, & Repp, 2007). This way, one would be able to generate predictions for one's own actions at the same time as predicting the timing of others' actions. A finding that may be best explained this way is that pianists duet better

(operationalized as the temporal difference between notes of the two different parts that should be played synchronously) when they play along with an earlier recording of themselves than when they are synchronizing with a part that stems from another pianist's earlier performance (Keller et al., 2007). So the assumption here is that one's predictions of perceived actions are more accurate for earlier self-generated actions, facilitating synchronization. (It should be noted, though, that this study does not rule out the possibility that pianists better synchronized with themselves because of the inherently greater similarity of the motor system.)

On the other hand, a more simple way in which the timing of others' actions could be (indirectly) taken into account is by focusing on joint effects of one's own and others' actions (Knoblich & Jordan, 2003). Rather than simulating another's part, one may simply generate predictions in relation to one's own performance and the goal to be achieved, thereby treating the others' actions as a kind of ''distortion.'' To illustrate, when learning to windsurf, one must learn to balance on the board and hold the sail so as to account for possible sudden changes in wind force and in the waves. One does not need to learn about specific effects of wind or waves separately, rather one will just try to control one's body so that the overall goal (keeping the sail up and not falling into the water) is met. Likewise, when acting with another person, one may learn to adjust the timing of one's own actions so that they fit with the others' by focusing on the consequences of one's combined efforts.

Thus, rather than acquiring a specific model about the other's performance, one might acquire a model for the joint performance. We believe that the acquisition of such joint internal models is quite likely to underlie joint learning. There is a lot of evidence suggesting that actions are planned in terms of the effects one wishes to accomplish (Prinz, 1997) rather than in more proximal terms (e.g., the exact movements one needs to make). The effects that are to be brought about through the coordination of one's own and the other's actions could provide the ''planning horizon'' against which outcomes of joint actions are compared. This account also makes a specific prediction as to how errors during joint action performance should be processed. Errors should weigh heavier if they occur at the level of the intended joint effects than if they occur at the level of individual performance. For example, if one of the musicians in a group hits a wrong note, he or she might not so easily become aware of it, provided the overall sound is the way it should be.

So far, there is little empirical work addressing the question of how temporal coordination in joint action is acquired. One exception is a study where individuals tried to learn to jointly control a distal event with or without receiving feedback about each other's actions (Knoblich & Jordan, 2003). The task was to keep a circle on top of a target moving horizontally along the computer screen, using an ''acceleration'' and a ''deceleration'' key. It was investigated how well single individuals could learn this task using both keys, compared to two participants in charge of one key each. Participants in the joint condition could not see each other, and it was manipulated whether they received auditory feedback about each other's actions (a tone for each key press).

The results showed that with practice, joint task performance reached the level of individual performance, but only when participants received feedback about the timing of each

other's actions. This suggests that in order to build up models of joint motor control, one must be able to link back the perceived consequences of joint actions to the actions of self and other. Without knowing how one's own and the other's actions relate in time, it is harder to learn to predict the consequences of the combined actions, and thus, the joint performance will never reach the same level of temporal coordination as the individual performance.

Unfortunately, there is not enough empirical evidence at present to further specify the role of parallel simulations and models of joint performance. It could be that these processes occur in different situations or at different stages of learning, and it is conceivable that they work in concert. We would like to stress, however, that in the absence of feedback regarding the timing of others' actions, developing models of joint performance seems like the only way to achieve coordination.

5. Where: Sharing space and attention

Joint actions occur in a shared physical space. This raises a number of questions about how co-actors manage this common space, ranging from how they avoid bumping into one another, to knowing that they are attending to the same object. In line with our overall theme of exploring the role of common coding in joint action, we will focus on two aspects: predicting spatial characteristics of others' actions (Erlhagen & Jancke, 2004; Jordan & Hunsinger, 2008) and sharing attention with others (Tomasello et al., 2005).

Sharing the same space means that actors must make predictions both about where the other is moving and about where objects the other acts upon will end up. This may go together, such as when the other is moving around a table to arrange objects. However, certain kinds of actions, such as throwing, may require predictions about the movement of body parts as well as the objects manipulated or propelled by them. In order for action simulation to be useful for joint action it ought to support such spatial predictions. Surprisingly, spatial predictions have rarely been explicitly addressed in empirical research on action perception and joint action (except maybe research into the so-called representational momentum effect; for references, see Wilson & Knoblich, 2005). This could be partly due to the fact that spatial predictions fall into the domain of mainstream perception research and that it is very hard to demonstrate that the action system makes an additional contribution.

5.1. Space

Conceptually speaking, it seems likely that simulating another's action can help us predict where the other is moving. In particular, spatial predictions could be a by-product of predictions about others' action goals. If we see others reaching for a heavy object on a shelf, we may predict at which locations their different body parts are going to be. Likewise, if a dancer sees his partner initiating a particular jump, he may be able to generate a prediction of the other's landing position based on an understanding of the type of jump the other

will perform. However, it may also be possible that movement trajectories can be directly predicted based on simulating observed actions, rather than coming just as a by-product. Just as real-time simulations can be used to generate predictions about others' timing, they may also be used to predict the other's movement path. An open question, however, is how exactly differences in perspectives could be overcome (e.g., Oztop, Kawato, & Arbib, 2006).

A related issue concerns predicting where objects will end up once the other has set them in motion through his or her actions—for instance, kicked, thrown, or pushed them. Findings from a study where people predicted the landing position of darts thrown at a target board support the idea that action simulation supports predictions about the end location of objects (Knoblich & Flach, 2001). It was shown that, having seen only the initial part of the throw, participants could predict the landing position more accurately when they saw a video recording of their own earlier actions than when they saw a recording of another person. This can be explained by the assumption that the predictions recruited the observers' own motor system. The closer the match between the system perceiving the action and the system performing the action, the more accurate the resulting predictions about the object location will be (cf. the logic of the self-recognition studies in the previous section). To summarize, action simulation may help one to determine where an object handled by the other will be in a few moments. This is, of course, highly useful for joint action coordination.

5.2. Attention

Joint action also requires attending to objects and events together. This second spatial demand is closely related to the ability to engage in joint attention. This ability has mostly been studied from a developmental and animal cognition perspective so far (e.g., Eilan, Hoerl, McCormack, & Roessler, 2004). Joint attention aligns the perceptions of different actors, helping to establish ''perceptual common ground,'' and thus reducing the work language has to do (Clark, 1996; Clark & Krych, 2004). This may increase the likelihood of the same action opportunities popping out, the same threats being detected, and the same knowledge getting activated (Sebanz & Knoblich, 2008).

There is general consent that joint attention is more than mere gaze following, because it involves awareness of perceptual experiences being shared (Call, this issue, pp. 368–379; Carpenter, this issue, pp. 380–391). So, it implies that both actors know that they are both attending to the same object or location. The cognitive and neural mechanisms underlying joint attention are still underspecified (but see Williams, Waiter, Perra, Perrett, & Whiten, 2005). It seems clear, though, that in order to have a shared understanding, one needs to understand that the relation between another person and an object is essentially the same as the relation between oneself and this object. The triadic relationship between self, other, and objects, seems to imply an agentic understanding of the other and the ability to keep self and other apart (Knoblich & Sebanz, 2008; Sebanz & Knoblich, 2008).

As noted above, action simulation does not involve the same level of social understanding, because it does not presuppose a self–other distinction. Nevertheless, it can be fruitful

to explore the question of whether our ability to simulate and predict others, actions contributes to joint attention. A first potential link arises from the observation that eye movements constitute actions. Just as the goal of a grasping action is to get hold of an object, the goal of an eye movement is to acquire perceptual information. This suggests that just as we can predict the outcome of a perceived grasping action, we may also be able to predict the outcome of perceived eye movements (Pierno et al., 2006). This could contribute to the ability to align one's own and the other's perception. This idea remains to be further explored in empirical studies.

A second link between action simulation and joint attention is suggested by the finding that the gaze behavior of participants performing action sequences and the gaze behavior of participants observing performance of the same action sequences is highly similar (Flanagan & Johansson, 2003). In this study, both actors and observers directed their gaze at objects shortly before they were to be grasped. This overlap can be explained by the assumption that observers' simulation of the perceived actions entailed the eye–hand coordination programs normally involved in performing the action. This indicates that simulating the instrumental part of an object-directed action (such as the hand movement used to grasp an object) might lead one to automatically align one's attention with the observed actor's.

Findings on inhibition of return across people also support this idea (Welsh et al., 2005). Inhibition of return refers to the finding that people are slower to detect a target when it appears in the same location as a stimulus presented just before. It has been speculated that this phenomenon has developed through evolution to optimize visual search behavior. The idea is that it makes more sense to look for something in a new location rather than returning to the same location over and over. Welsh et al. (2005) demonstrated that inhibition of return also guides the search behavior of two people performing a task together. When one participant responded to a target in a particular location, the other participant was subsequently slower in detecting a target appearing at the same location. This finding can be explained by the assumption that each individual simulated the other's action, thereby shifting their attention to the other's target as if they were to respond to it. This then triggered the same inhibitory neural processes as when participants responded to a target themselves (Welsh et al., 2005), lending further support to the idea that through action simulation, people's focus of attention can become aligned.

6. The highs and lows of joint action

We hope to have shown that action simulation is a powerful mechanism for predicting the what, where, and when of others' actions. Although common coding of perceived and performed actions alone does not enable one to engage in joint action, common coding provides a representational platform for integrating the actions of self and other. More precisely, common codes allow actors to flexibly run and link different action simulations pertaining to self and other. At the same time higher-level task representations can separate one's own and others' contributions to a joint action.

We would like to end with some comments on potential links between action-related processing as discussed here and the thinking processes cognitive scientists have previously invoked to explain joint action—in particular, language and mental state attribution. Historically, philosophers of action first highlighted joint action as an interesting topic of scientific study. The focus was on joint intentionality (cf. Tollefsen, 2005) and on the question of how people manage to agree on doing things together, neglecting a bit how people actually get things done. Of course, talking and thinking are very powerful when it comes to making pacts and negotiating action plans, but we believe that the fine-grained spatial and temporal coordination needed to get things done is a different matter.

To be sure, work on discourse has revealed many ways in which language is used as a coordination device (Brennan, this issue, pp. 274–291; Clark, 1996) and has paved the way for current joint action research. However, the focus was on understanding of how language could be conceived as joint action and not on the processes supporting nonverbal joint action. One main challenge for future work seems to be to understand how lower-level processes like action simulation and higher-level processes like verbal communication and mental state attribution work in concert, and under which circumstances they can overrule each other. Before we get to understand these interactions, a good dose of ecological, motor, and evolutionary spices seems like a healthy addition to the heavily mentalistic stew previously cooked up in solitary armchairs.

References

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

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology. Human Experimental Psychology*, *53*, 153–164.

Blakemore, S., Frith, C., & Wolpert, D. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, *11*, 551–559.

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

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62–81.

Davidson, P., & Wolpert, D.M. (2003). Motor learning and prediction in a variable environment. *Current Opinion in Neurobiology*, *13*, 1–6.

Eilan, N., Hoerl, C., McCormack, T., & Roessler, J. (2004). *Joint attention: Communication and other minds*. Oxford, England: Oxford University Press.

Erlhagen, W., & Bicho, E. (2006). The dynamic neural field approach to cognitive robotics. *Journal of Neural Engineering*, *3*, R36–R54.

Erlhagen, W., & Jancke, D. (2004). The role of action plans and cognitive factors in motion extrapolation: A modelling study. *Visual Cognition*, *11*, 315–341.

Fetzer A (Ed.) (2007). *Context and appropriateness: Micro meets macro*. Amsterdam: John Benjamins Publishing Company.

Flach, R., Knoblich, G., & Prinz, W. (2004). Recognizing one's own clapping: The role of temporal cues in self-recognition. *Psychological Research*, *11*, 147–156.

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*, 769–771.

Grafton, S., & Hamilton, A. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*.

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, *77*, 73–99.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge, England: Cambridge University Press.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology*, *12*, 123–135.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, *19*, 265–288.

Huygens, C. (1665). *Kort onderwijs aengaende het gebruyck der Horologien tot het vinden der Lenghten van Oost en West*. The Hague.

James, W. (1890). *The principles of psychology* (Vol. 2). New York: Holt.

Jordan, J. S., & Hunsinger, M. (2008). Learned patterns of action effect anticipation contribute to the spatial displacement of continuously moving stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 113–124.

Keller, P. (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., Knoblich, G., & Repp, B. (2007). Pianists duet better when they play with themselves. *Consciousness and Cognition*, *16*, 102–111.

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S.-J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, *7*, 1299–1301.

Knoblich, G., & Flach, R. (2001). Predicting action effects: Interactions between perception and action. *Psychological Science*, *12*, 467–472.

Knoblich, G., & Jordan, S. (2002). The mirror system and joint action. In M. I. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 115–124). Amsterdam: John Benjamins.

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

Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 456–465.

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: From entrainment to joint action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*, 2021–2031.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*, 846–848.

Lakin, J., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, *14*, 334–339.

Lotze, H. (1852). *Medicinische Psychologie und oder Physiologie der Seele*. Leipzig: Weidmannsche Buchhandlung.

Neda, Z., Ravasz, E., Brechte, Y., Vicsek, T., & Barabasi, A.-L. (2000). The sound of many hands clapping. *Nature*, *403*, 849–850.

Ouillier, O., DeGuzman, G. C., Jamtzen, K. J., Lagarde, J., & Kelso, J. A. S. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience*, *3*, 178–192.

Oztop, E., Kawato, M., & Arbib, M. A. (2006). Mirror neurons and imitation: A computationally guided review. *Neural Networks*, *19*, 254–271.

Pacherie, E., & Dokic, J. (2006). From mirror neurons to joint actions. *Cognitive Systems Research*, *7*, 101–112.

Pierno, A. C., Becchio, C., Wall, M. B., Smith, A. T., Turella, L., & Castiello, U. (2006). When gaze turns into grasp. *Journal of Cognitive Neuroscience*, *18*, 2130–2137.

Prinz, W. (1990). A common-coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action: Current approaches* (pp. 167–201). Berlin: Springer-Verlag.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*, 129–154.

Ramnani, N., & Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nature Neuroscience*, *7*, 85–90.

Repp, B. H., & Knoblich, G. (2004). Perceiving action identity: How pianists recognize their own performances. *Psychological Science*, *15*, 604–609.

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

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Reviews of Neuroscience*, *27*, 169–192.

Roepstorff, A., & Frith, C. (2004). What's at the top in the top-down control of action? Script-sharing and 'top-top' control of action in cognitive experiments. *Psychological Research*, *68*, 189–198.

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Consciousness and Cognition*, *17*, 1219–1227.

Schmidt, R. C., Bienvenu, M., Fitzpatrick, P. A., & Amazeen, P. G. (1998). A comparison of intra- and interpersonal interlimb coordination: Coordination breakdowns and coupling strength. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 884–900.

Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: Action-induced modulation of perception. *Trends in Cognitive Sciences*, *11*, 349–355.

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

Sebanz, N., & Knoblich, G. (2008). From mirroring to joint action. In I. Wachsmuth, M. Lenzen & G. Knoblich (Eds.), *Embodied communication* (pp. 129–150). Oxford, England: Oxford University Press.

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. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1234–1246.

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

Sebanz, N., Rebbechi, D., Knoblich, G., Prinz, W., & Frith, C. (2007). Is it really my turn? An event-related fMRI study of task sharing. *Social Neuroscience*, *2*, 81–95.

Stanley, J., Gowen, E., & Miall, C. (2007). Effects of agency on movement interference during observation of a moving dot stimulus. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 915–926.

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

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.

Vygotsky, L.S. (1987). *The collected works of L.S. Vygotsky (Vol.1): Problems of general psychology*. Including the volume Thinking and Speech. New York: Plenum.

Watanabe, K. (2008). Behavioral speed contagion: Automatic modulation of movement timing by observation of body movements. *Cognition*, *106*, 1514–1524.

Welsh, T. N., Elliott, D., Anson, J. G., Dhillon, V., Weeks, D. J., Lyons, J. L., & Chua, R. (2005). Does Joe influence Fred's action? Inhibition of return across different nervous systems. *Neuroscience Letters*, *385*, 99–104.

Williams, J. H. G., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *Neuroimage*, *25*, 133–140.

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