A framework for joint action

The core characteristic of the framework for joint action, as depicted in Fig. 1, is the differentiation between the individual herself (ME), the interaction partner (YOU) and the joint aspect (WE). This distinction originates from philosophical accounts highlighting the importance of shared intentions for joint action (e.g., Bratman, Searle, Tuomela). Common to these approaches is the idea that, in order to perform a true joint action (i.e. one that does not simply have an accidental coordinated outcome), people have to *share the* *intention* to perform the action jointly – including all the consequent mechanisms such as behaving cooperatively towards the other, forming joint representations, predicting and monitoring the partner’s behavior and adjusting one´s own behavior in a way that leads to the probable success of the shared task (Sebanz, Bekkering, & Knoblich, 2006). Thus, joint action is more than performing individual actions at the same time and space (if at all) because it explicitly requires that there are intentions, representations and processes on the level of the group or the dyad that go further than what individuals intend or represent about their own action parts.

The accounts on shared intentionality thus influenced our thinking about different representational levels involved in joint action because they emphasize the role of the WE – that what presumably differentiates between individual and joint actions. However, besides what we know from philosophers and, in part, from developmental psychologists and linguists who further developed these ideas (e.g., Tomasello, Brennan & Clark), little is known about the WE-aspect in processes involved in actual joint action performance. In other words, an explicit notion of a WE has so far only been formulated with respect to representations, but not so much for the lower levels of action such as simulation and monitoring processes or action strategies. This is unfortunate because representations as such do not explain how people actually manage to coordinate their actions online.

In the framework suggested in this paper (Fig. 1), we link existing evidence on ME-, YOU- and WE-distinctions with considerations of what questions need to be asked in more detail in the future. Our core suggestion, here, is that also the mechanisms and processes underlying joint action have separate ME-, YOU- and WE-aspects. That means that not only representations are different in individual action and joint action, but also the mechanisms and processes required for planning and executing joint actions differ from those required for individual action. (Please note, however, that this differentiation is a functional one: At this point, we do not mean to say anything about anatomical similarities or differences, i.e. where in the brain these aspects are processed, or about causal relations and hierarchies, i.e. whether ME, YOU and WE are independent of each other and how they relate to each other.)

The framework focusses on two action processes that have frequently been shown to be important for any type of action: Simulation, on the one hand, is the ability to internally run simulations of possible actions and their outcomes. This helps to predict the likely consequences of own actions (Wolpert, Repp\_PsycSci\_2004, Repp\_PsychSci\_2004, Keller\_ConscCog\_2007) as well as of other people’s actions that are observed (Sebanz & Knoblich, 2009; Wilson & Knoblich, 2005). Being such a universal phenomenon, it seems plausible that it is also involved in the most advanced form of social interaction, i.e., the joint case. Although evidence on this aspect is sparse, there is currently an important discussion about the nature of the simulation processes involved in joint action (Keller). The crucial question is in what way simulation of an observed other’s action and simulation of an interaction partner’s action differs: If joint action is more than just the individual action parts performed next to each other, then the simulation process needs to take this joint aspect into account. We will come back to this still open question in the section on empirical evidence.

Monitoring, on the other hand, is the process of observing the state of one’s actions and the context they occur in. This is important, for example, to detect errors in own (xxx) and observed others’ actions (de Bruijn). It is also useful for flexibly adapting one’s own behavior according to what an interaction partner is doing in a joint context although so far only little is known about monitoring processes in joint action. An open question, for example, is whether individuals monitor only their own and their partner’s actions or whether there is a joint monitoring aspect involved. In specific, is the joint action performance monitored independently of the individual action parts? We will also return to this question in detail in the empirical evidence section.

A further innovation of our framework for joint action is that it provides a structure for discussing the effects of strategies. Specifically, we claim that people might not always make use of all components depicted in Fig. 1 – be it because it is not possible given the available information or because it is not necessary for solving the joint task. Thus, by emphasizing or reducing the impact of certain components on one’s own action planning and execution (indicated by the curved arrows in Fig. 2), joint action tasks might be solved in different, possibly more efficient ways. As an intiuitive example think of the situation in which you and a friend have to meet in Paris tomorrow without being able to communicate the exact place and time (“Schelling game”, Clark, 1996). What you lack in this case is precise information about what the other is doing; we can say that there is no online monitoring of the partner possible. Thus, in this case you have to find a strategy that can solve your joint coordination problem without making use of this particular mechanism that one might normally rely on during joint action tasks. (A solution to this problem is to take into account the common ground of the interacting persons (Clark, 1996; more) and to choose a time and place that is most salient for you and your friend, e.g., at 12 am on top of the Eiffel tower.) The idea, here, is that in most joint action tasks, people may make use of such strategies, thereby focussing on certain aspects, e.g., more on their own task than on the other’s or mainly on the joint outcome etc. Strategies seem especially important for cases in which precise timing of the individual action parts is required (Vesper, in prep.). We will elaborate on this idea in the section about strategies.