Acta Psychologica 215 (2021) 103285

Contents lists available at ScienceDirect

Acta Psychologica

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

Editorial

# Understanding joint action: Current theoretical and empirical approaches

ARTICLE INFO

*Keywords*
Joint action

ABSTRACT

Joint actions are omnipresent, ranging from a handshake between two people to the coordination of groups of people playing in an orchestra. We are highly skilled at coordinating our actions with those of others to reach common goals and rely on this ability throughout our daily lives. What are the social, cognitive and neural processes underlying this ability? How do others around us influence our task representations? How does joint action influence interpersonal interactions? How do language and gesture support joint action? What differentiates joint action from individual action?

This article forms an introductory editorial to the field of joint action. It accompanies contributions to the special issue entitled "Current Issues in Joint Action Research". The issue brings together conceptual and empirical approaches on different topics, ranging from lower-level issues such as the link between perception and joint action, to higher-level issues such as language as a form of joint action.

## 1. Introduction

The term "joint action" refers to actions in which two or more people coordinate in space and time in order to bring about a change in the environment (N. Sebanz, Bekkering, & Knoblich, 2006). Research on joint actions includes coordination of both verbal and non-verbal interactions. Whereas research on verbal joint action has a long history in the context of language and communication, the focus on non-verbal forms of joint action largely emerged over the past two decades.

In this article, we will provide an overview of approaches to joint action research and integrate current research. This article accompanies a special issue entitled *Current Issues in Joint Action Research*. Both this article and the special issue are meant to provide a starting point for anyone interested in the study of joint action, as well as to those already in that field. We do so by first situating joint action research and by briefly providing an overview of its origins. We then introduce some of the main approaches and topics, while integrating the articles that are part of the special issue. As the articles do not cover all of the topics of interest to the joint action community, we will end this article by providing pointers to several additional topics related to joint action as well as some future challenges.

### *1.1. The field of joint action research*

The field of joint action research falls at the intersection of cognitive science and social psychology, while also drawing interest from philosophers, developmental psychologists, roboticists, and other related fields. The field originated in part from the realization that, while there had been a substantial number of studies on the processing of social information by individual participants reacting to stimuli on a computer screen, very little work had investigated the mechanisms involved in more (often literally) hands-on social interaction. This was particularly true for social interactions involving the planning and execution of non-verbal forms of joint action.

To remedy this observation, one of the first approaches was to put individual cognition and coordination into socially embedded scenarios to study the interplay between basic cognitive mechanisms and social contexts. In a paradigmatic study, N. Sebanz, Knoblich, and Prinz (2005) used a classic Simon response-compatibility task (J.R. Simon, 1990; J.R. Simon, Hinrichs, & Craft, 1970) embedded in a social context. In the task, participants saw a stimulus image of a pointing index finger that had a ring placed on it. The ring was either green or red and participants responded to the color of a ring by pressing one of two buttons for a green ring and another for a red ring with one of their index fingers. The pointing finger with the ring on it happened to either point towards the response location (in compatible trials) or to the other response location (in incompatible trials). When individuals completed this task by themselves, they showed a standard Simon effect; they responded more quickly on the compatible compared to the incompatible trials. In a control condition, they completed this same task while the other half of the task was left open (i.e., a Go/No-go task in which there only was a response when the ring color was assigned to the participant). They did not show a Simon effect in this individual version of the task. The novel twist concerned a joint condition, in which two participants completed the task together while each only responded to one of the colors with one of the button presses. Thus, the task in both the joint and the individual control condition was identical, except that the no-go part of the task would require a response by the other participant in the joint condition. Surprisingly, the Simon effect re-emerged in this joint task. This finding is known as the joint Simon effect.

https://doi.org/10.1016/j.actpsy.2021.103285

Available online 3 March 2021

The joint Simon effect gave rise to the notion of task co-representation, suggesting that people form representations of co-actors' tasks performed around them. There is some debate about how task co-representation exactly works (see W. Prinz, 2015), but it is clear that other actors form a salient feature in task contexts and that the presence of other actors and the specifics of their tasks may influence our own task performance. This general observation has been demonstrated across a range of tasks. Some of these tasks do not involve choosing between responses along a spatial stimulus dimension, but reveal task co-representation in non-spatial domains, such as in studies on joint attention (Böckler, Knoblich, & Sebanz, 2012), music making (Loehr, Kourtis, Vesper, Sebanz, & Knoblich, 2013) and memory performance (Eskenazi, Doerrfeld, Logan, Knoblich, & Sebanz, 2018), to name a few. Regardless of the exact mechanics of task co-representation, it is clear that co-representation accounts are representational in nature. By this, we mean that these accounts use internal codes as stand-ins for external states and events. As intermediaries between perception and action, these codes tend to be symbolic and computational in nature, and serve the purpose of directing behavior. Representations can be flexibly activated without needing the immediate presence of what it is they represent (e.g., Haugeland, 1991).

Another paradigmatic set of early studies on joint action illustrates a different approach. This approach can be referred to as the dynamical systems approach (which, like the term representational approach, is an umbrella term for a range of more specific approaches). It differs from the representational approach in both its theoretical view as well as in the types of tasks used to study joint action. With respect to the main theoretical difference, dynamical systems approaches do not rely on representations in the way we described above. Instead, they postulate that behavior emerges from the interplay between organisms and their environment. With respect to the tasks used to study joint action, dynamical systems studies typically involve some form of continuous movement, such as walking or swinging a pendulum. Rather than giving participants a discrete task in which they need to provide specific responses to specific stimuli, these studies aim to understand behavior as it unfolds over time. This approach is extremely powerful for understanding continuous joint coordination, be it between schools of fish, flocks of birds, or interpersonal human coordination.

A wealth of dynamical systems studies on joint action demonstrate a core mechanism for coordination, namely entrainment. Entrainment refers to the tendency for spatiotemporal coordination to occur spontaneously between two parts of a moving system. These parts are not necessarily physically coupled, as entrainment occurs across people as well. For example, when rocking in rocking chairs alongside each other, participants tended to entrain, such that they moved in synchrony more often than would be expected by chance. This was even the case when participants rocked in rocking chairs that had different natural frequencies (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007). People also entrain when they engage in conversation by synchronizing their body sway (Shockley, Santana, & Fowler, 2003) and when audiences clap in unison for a theatre or opera performance (Néda, Ravasz, Vicsek, Brechet, & Barabási, 2000). Thus, synchronization is a widespread phenomenon in interpersonal coordination, and it may support joint action.

## 2. Contributions to the special issue

Aside from using a variety of approaches and a variety of tasks, joint action research also considers a variety of different processing levels. They include low-level sensory coordination, perception-action coupling, joint attention, joint action planning, language and communication, the experience of acting jointly, and implementation in the brain and in artificial systems, to name a few. The diversity in levels also brings along a variety of measures, ranging from qualitative and experiential measures, to performance measures such as reaction times and accuracy, to measures of neural activity. The collection of articles in this special issue reflects some of this diversity as well.

The first set of articles we will discuss provides examples of recent representational approaches to joint attention, perspective taking, and learning and memory in joint settings. We will then discuss articles on the effects of interpersonal coordination on prosocial behavior. Then, we move on to some examples of joint action in more complex task settings, such as joint communication and coordination of tones in musical ensembles. We will also consider joint task distributions in this section. We then turn towards the implementation of joint action mechanisms in biological and artificial systems. Finally, we will discuss some other areas of joint action, as well as remaining challenges for the field.

### 2.1. *Contributions to representational approaches*

One issue with respect to how actions are represented concerns the overlap in perceptual and action codes. This issue has been formalized in theories such as the Common Coding theory (W. Prinz, 1990) and the Theory of Event Coding (Hommel, Müsseler, Aschersleben, & Prinz, 2002). The overlap of perception and action is of interest to joint action researchers, as it may provide a low-level mechanism based on which our own and other's actions can be conceived of in commensurate units. This would then support integrating one's own and others' actions during joint actions.

One way to study how seeing actions by others around us may influence our own social-cognitive processing is by virtue of cueing studies. In these studies, cues such as eye-gaze or pointed fingers may be used to determine to what extent these body parts may direct our attention towards particular locations. It has been shown that eye-gaze and pointed fingers increase the salience of the pointed-to locations (a social cueing effect). In a study by T.N. Welsh, Reid, Manson, Constable, and Tremblay (2020) (in this special issue), they examined how hand and foot cues could similarly direct attention, and whether the extent to which they did depended on the body-part used to provide the response to a stimulus. To do so, they provided non-informative directional hand or foot cues while participants responded to targets that could show up in one of two locations shortly after the cues. As is the case in most cueing paradigms, the cue could either point towards the target location (congruent trials) or to the other location (incongruent trials). The results indicated that, while hand cues showed a compatibility effect for responses made with either the hands or feet, foot cues did not. The authors attributed this difference to the possibility that hands tend to have more social relevance in understanding others' actions than feet do.

In another cueing study, Gobel and Giesbrecht (2020) (in this special issue) sought to determine how joint attention would modulate performance in a target detection task. In the task, a participant and a confederate saw a cue that participants were led to believe to indicate the gaze location of the confederate. The target would then appear in the cued location or in an uncued location. In the covert attention experiment, participants were told not to move their eyes, whereas in the overt attention version, they could. The authors also varied whether the confederate was lower or higher social rank, by manipulating the information they gave to the participants about their task partner. Their findings indicated that participants showed an inhibition of return effect, meaning that they were faster for the cued locations than the uncued locations when the duration between the cue and the stimulus (SOA) was 150 milliseconds. For longer durations, the opposite effect emerged. Interestingly, this inhibition of return effect was stronger for the higher social rank than the lower social rank group. This modulation only emerged for the overt attention experiment and not for the covert attention experiment. The authors argue that overt attention is thus more susceptible to social factors.

In another joint attention study that focused on social factors, Nafcha, Morshed-Sakran, Shamay-Tsoory, and Gabay (2020) (in this special issue) examined whether the previously established Social Inhibition of Return (SIOR effect; T.N. Welsh et al., 2005) would be modulated by group membership. The SIOR effect refers to the finding that people are slower to detect a target at a location already attended to by another agent than at a new location. The authors tested the prediction that the SIOR effect would be stronger for pairs of in-group members compared to out-group members. To do so, the authors conducted this study in Israel and used a pre-existing in-group and out-group scenario by pairing Jewish and Muslim students. In line with their predictions, they found a stronger SIOR effect for in-group than for out-group members, suggesting that social factors may influence basic reflexive cognitive processes. These results are also in line with the proposal that overt attention may be susceptible to social factors (Gobel & Giesbrecht, 2020), as this paradigm involved overt attention as well.

The notion that joint processing may differ for tasks involving overt versus covert attention aligns with a study by Constable and Knoblich (2020) (in this special issue) as well. In their task, they examined whether a previously reported self-prioritization effect would differentially facilitate or hinder responses to stimuli that were previously associated with a partner or a stranger. Participants responded to a set of shapes, each of which was assigned to themselves (the self), to a partner, or to a stranger during a training phase. During the matching task phase, participants then saw a shape and an identity label (i.e., a name) presented above and below a fixation cross, respectively. They had to respond as quickly and accurately as possible by pressing one of two buttons, one in case of a match and one in case of a mismatch. Although there was a fixation cross, participants were allowed to move their eyes during this task. After a first block of trials, the pairings between shapes and names changed while participants performed a second training phase. They then repeated the matching task.

The results indicated a clear self-prioritization effect, as participants responded more quickly to their own shape and identity than to those of the partner or stranger. The results did not indicate systematic differences between the partner's or the stranger's shapes. Interestingly, reshuffling the pairings modulated response times. In particular, participants responded more slowly to shapes that previously belonged to their partner but now belonged to the self. This effect did not occur for shapes that previously belonged to a stranger. Based on these findings, the authors argue that information that was previously bound to a relevant task partner may be difficult to re-bind to the self. Thus, representational processes may depend on both a task partner's identity as well as on task history.

### 2.2. Contributions to interpersonal coordination

The studies just discussed provide a glance into some of the work on representational processes supporting joint action. These tasks share with each other that they require specific discrete responses (e.g., button presses) to particular stimuli. We will now discuss some of the work on interpersonal coordination, which tends to take a dynamical systems approach and employ more continuous action tasks (see R.P.R.D. van der Wel & Fu, 2015), for a discussion of the significance of these task differences). In dynamical systems studies on joint coordination, participants often produce some sort of continuous data input, such as when participants walk, swing pendulums, talk, listen, or dance. The data are then subjected to quantitative methodologies used to analyze time series, such as cross-recurrence quantification analysis, detrended fluctuation analysis, or spectral analysis. The goal behind these analyses is to understand systematic fluctuations in the data that emerge over extended time periods.

As an example of this approach, Gvirts Problovski et al. (2021) (in this special issue) investigate whether diminished abilities for interpersonal coordination could explain deficiencies in social cognition. More specifically, they postulate that this might be the case in adults with Attention Deficit Hyperactivity Disorder (ADHD). To test this hypothesis Gvirts Problovski et al. (2021) compare a sample of participants with ADHD with a control group in terms of abilities to intentionally or spontaneously synchronize with an experimenter. A Leap Motion controller recorded hand and arm movements during the task. These movements were then quantified by calculating a cosine velocity vector, which reflected the similarity in the movement profiles of the experimenter and participant. Interestingly, only in the intentional synchronization condition there was a significant difference in movement synchronization between the two groups. These results add support for the hypothesis that problems with interpersonal coordination are linked to disorders that include deficits in social cognition.

Dynamical systems approaches to joint action, such as the studies just described, have produced a wealth of data on possible mechanisms for interpersonal coordination. We encourage the reader to learn more about this approach (e.g., Schmidt, Fitzpatrick, Caron, & Mergeche, 2011), as there is a substantial literature on interpersonal coordination and their coordination dynamics. There has also been a collection of studies that focus on prosocial behavior or attitudes that are elicited by interpersonal coordination. The term "prosocial behavior" refers to voluntary actions that are intended to help or benefit another individual or group of individuals (Eisenberg, 1982). Among others, these behaviors include helping behaviors, cooperation, and comforting. We will discuss some recent examples of such studies in the next section.

### 2.3. Contributions to interpersonal coordination and prosocial behavior

In the recent decade work on interpersonal temporal coordination has produced a wealth of data that links various types of coordination with changes in prosocial behavior or attitudes. However, an underlying theory that can account for these findings has yet to emerge. Michael, McEllin, and Felber (2020) (in this special issue) develop a theoretical approach to systematically categorize different forms of interpersonal coordination. Starting from a minimal definition of coordination allows the authors to define distinct features of coordination, such as whether coordination is bidirectional or unidirectional and whether coordination is achieved via a direct link among partners or mediated by an external agent or entity. The authors also consider both emergent and planned coordination as well as instances in which coordination is only attempted without being successful. Finally, a distinction is made between coordination of actions and of decisions. Combining these distinctions with a revised definition of possible prosocial effects allows Michael et al. (2020) to spell out several possible mechanisms, both on an individual level as well as on a group-level. The first group of proposed mechanisms are based on the possibility that coordination may affect sources of motivation to engage in prosocial behavior. These sources include concern for others' well-being, trust, and a sense of commitment. The second group of mechanisms is based on the possibility that coordination may help to identify another person as the appropriate target of prosocial behavior. The authors provide possible manipulations and predictions, making it a good starting point for future research on the link between coordination and prosocial effects.

Atherton et al. (in this special issue) propose another possible underlying mechanism for the link between coordination and prosocial behavior, namely self-construal, and test it in a mixed-method study. The term self-construal refers to the extent to which a person is defined independently or interdependently of others. The hypothesis is that interpersonal coordination leads to a more interdependent self-construal, which in turn leads to more prosocial behavior. This hypothesis can be seen as an extension to the range of mechanisms proposed by Michael et al. (2020) (this special issue). While Michael et al. (2020) postulate that coordination might help to identify appropriate targets of prosocial behavior, Atherton et al. hypothesize that interdependent self-construal would lower the general threshold of identifying others as appropriate targets for prosocial behavior. Atherton et al. invited participants to engage in a joystick movement task, while a confederate's movements were either coordinated or uncoordinated with the participant's movements. Following this manipulation, participants filled in the Twenty Statements Test (TST) by (Kuhn & McPartland, 1954), in which participants have to generate twenty statements to the open-ended question “Who am I?”. Responses to the TST were then categorized as independent or interdependent. Examples for the latter include “I am a sister” and “I am a Christian”. The data show that participants in the coordinated movement condition reported a significantly larger proportion of interdependent responses than participants in the uncoordinated condition. A more detailed thematic analysis of responses further suggests that the difference between participants from the two movement conditions was especially pronounced for broader social constructs such as “Christian”, “African”, “Working class”, but not for constructs related to close social relationships such as family and friends.

Prosocial effects of coordination are often studied right after a coordination manipulation took place. Cross, Michael, Wilsdon, Henson, and Atherton (2020) (in this special issue) extend our knowledge of these effects by investigating whether such effects persist in time. In Study 1, Cross et al. (2020) use the same movement manipulation as Atherton et al. (this special issue). They first completed a task in which they produced coordinated or un-coordinated actions with an experimenter. They were then asked to complete a survey sent to them 24 h after the study to help out the experimenter. Those who had produced coordinated movements were more likely to help the experimenter by completing the survey than those who produced un-coordinated movements. In Study 2, this persistence of helping behavior across 24 h was replicated with a tapping-based coordination task of shorter duration. The findings indicated that even short periods of coordinated interactions can enhance the prosocial behavior over an extended period of time.

Kato et al. (2020) (in this special issue) approach the question of prosocial effects of coordination by investigating how walking and talking impact interpersonal impressions. The authors consider multiple possible factors for the relation between coordination and its prosocial effects. They carefully specify a model of several possible paths of influence that includes variables such as prior contact, turn-taking in speech, speech overlap, synchronization of walking and leg length differences across participants. A 90-min-long social interaction period before participants walked with each other was used to manipulate prior contact, such that participants would either walk with someone they have already been interacting with for 90 min or someone from a different group with whom they have not had that interaction period. The data supports the hypotheses that leg length difference influences synchronization of walking which in turn influences interpersonal impressions. Interpersonal impressions were also influenced by prior contact, turn-taking in speech and speech overlap. A possible mutual influence of synchronization of walking and coordination of speech is not supported by the data and neither an influence of prior contact on walking nor on speech coordination. Taken together these findings suggest that walking and speech coordination independently impact interpersonal impressions in naturalistic strolling behavior, while interpersonal impressions did not show significant effects on coordination.

Taken together, these contributions to the special issue form a good starting point and viable resource for joint action researchers who are interested in the far reaching and lasting prosocial effects of interpersonal coordination. Before turning to more complex forms of coordination, we first turn from the effects of joint coordination on prosocial behavior to the perception of joint-ness when participants watch groups of people act with different levels of coordination. In particular, Lee, Launay, and Stewart (2020) (this special issue) nicely generated groups of seven virtual avatars that performed a set of dance routines. The authors implemented conditions in which the group either acted in unison (meaning that they all performed the same dance routine) or in coordination (meaning that three of the seven performed one routine while the other four performed a different routine). They also manipulated whether the performance was aligned or misaligned, by applying lag times to the misaligned conditions. Participants then watched the different movie clips and rated how socially bonded and how formidable (i.e., powerful or daunting) the group seemed to them. The results indicated that dancers were perceived to be more socially bonded when moving in unison and during temporally aligned coordination. The study provides an interesting example that speaks to the powerful nature of joint action, both in terms of the ability to produce effects one cannot produce alone (which is the case in synchronized dance performances) as well as its perception of such joint effects (as Lee et al., 2020 demonstrate).

Coordination among multiple members of a group is one example of a more complex joint action. In the next section, we provide some examples of other approaches to joint action that focus on more complex tasks. Doing so is particularly useful, as it is likely that many joint actions in real life involve a meshing of many processing levels and mechanisms. Those range from mechanisms postulated by representational and dynamical systems accounts, but also range from low-level sensory input to more high-level abstract processing (see Scott Jordan, Schloesser, Bai, & Abney, 2018, for an example discussion of such multiscale contingencies).

### *2.4. Contributions to communication and other complex joint actions*

In this section, we will provide several examples of how our sensorimotor systems can scaffold communication and coordination in more complex joint action settings. As a first example of a more complex joint action task, Schmitz, Knoblich, Deroy, and Vesper (2021) (in this special issue) investigate how perceptual and cognitive similarities between coactors can be used as a communication tool. They put forward the interesting proposal that cross-modal common ground, i.e. the cross-modal associations between perceptual features of objects that are available to interactive partners by virtue of our shared perceptual capacities, can serve as coordination and communication tools. In the context of a public engagement event in an art museum, pairs of museum visitors participated together in a communication game, in which one of them (the ‘sender’) informed their partner (the ‘receiver’) about different visual stimuli. Participants were asked to spontaneously create a novel communication system using one sensory modality as the referent (e.g., visual stimuli of different sizes to communicate about) and another as the communication medium (e.g., piano tones of different auditory pitch to communicate with). The authors show that participants reliably made use of cross-modal correspondences for communication and assumed that such signals would be easily understood by their task partner (i.e. assumed cross-modal common ground). They also show that, when participants were later exposed to ambiguous stimuli mappings, they relied on their interaction history to re-establish a communication system. These findings show how individuals flexibly navigate complex interactive scenarios by relying on shared perceptual and cognitive capacities, and on the shared history with their interactive partner.

Another interactive domain where individuals may rely on common ground and shared history in order to establish interpersonal coordination is dialogue. Lelonkiewicz and Gambi (2020) (in this special issue) investigate what coordination strategies speakers adopt in conversations to ensure coordination and alignment. They present a novel contribution that focuses on speakers’ rather than listeners’ coordination strategies and show how speakers increase the predictability of both the timing and the content of their words to foster interpersonal alignment and help listeners to coordinate. In a Word Chain Task, pairs of participants were given a mystery word and asked to produce a definition that would allow a third party to guess the mystery word. Participants interacted via a chat-based interface that allowed them to contribute one word at a time. In a control solo condition, a single participant was given the same mystery words and asked to produce a definition that would allow another person to guess it. Their results show that speakers reduced the variability of the inter-turn intervals (lag between consecutive words they produced), providing interesting evidence in support of the implementation of variability reduction as a coordination strategy in the domain of linguistic interactions. The findings of this study interestingly highlight that certain coordination strategies, such as variability reduction, can be implemented by interactive agents in different interactive domains. This suggests that similar mechanisms are at play when individuals work together towards a common goal, whether it concerns coordinating their actions in space and time, or successfully communicating relevant information to each other.

In another example of coordination during linguistic interactions, Hadley, Fisher, and Pickering (2020) (in this special issue) focuses on the role of predictive simulation in determining how well listeners can anticipate the timing of their conversational partners and prepare their timely response. Like many joint actions, conversation is a very complex task which can be performed with milliseconds level precision and relies on motor simulative processes supporting prediction and adaptation. This paper investigates the role of listeners' motor experience in determining the temporal prediction of turn-ends and utterance production in a turn-based conversation task. The authors propose that listeners not only predict the timing of the upcoming word, but also the style with which the word will be uttered. Because their predictions are partly based on their previous motor experience with the given word, the larger the discrepancy between their own and the speakers' style, the worse their predictions will be. The results of their study show that participants are faster and more accurate in responding to a speaker's production when the speaker's uttering style is similar to their own, and they are as accurate as when responding to a recording of their own utterances. This evidence supports the hypothesis that idiosyncratic motor experience plays an important role in our ability to accurately predict a partner during a complex joint action. These findings contribute to our understanding of the role of prediction, simulation and motor experience in complex joint actions. In complex joint actions, coordination is supported by predictive abilities on the receivers' side, and prediction is facilitated by similarity in movement and production styles between the sender and the receiver.

Another paradigmatic case of complex joint action is ensemble music making where partners, like in conversations, need to constantly predict, adjust and communicate with each other in order to align their productions at the millisecond scale. Six et al. (in this special issue) investigate the role of gestures and bodily movements in supporting interpersonal musical timing and role dynamics during a joint music making task. By means of a tapping pad, participants could jointly produce a melody: one participant had the upper voice, the other one the lower voice, and both together were equally contributing to the musical outcome. The authors measured the synchrony in timing between the two participants as a function of their mutual visual access, and of the presence of expressive or neutral gestures accompanying the tapping movements. The results of this study suggest that expressive gestures and mutual visual information exchange facilitate joint musical timing. These findings interestingly reveal how communication among co-actors can be effectively established through nonverbal channels, and importantly show how alignment in highly complex joint action scenarios depends also on subtle movement kinematic cues, such expressive gestures and ancillary movements that provide time-varying cues supporting anticipation and prediction.

In complex interactive contexts agents need to deal with an overwhelming amount of information coming from their partner and the constantly changing environment. In such situations, it becomes important to parse perceptual information and learn how to flexibly prioritize certain perceptual cues over others in order to support coordination. In a study investigating joint music making using *E*-music boxes, Liebermann-Jordanidis et al. (in this special issue), explored the role of self-other segregation and integration processes when dealing with both visual and auditory information coming from an interactive partner. Pairs of participants were asked to play a musical instrument, while the experimenters manipulated their parts so they would vary in terms of the frequency of movements required to play, and the pitch of the tones produced. Their results on interpersonal synchrony show that pairs of participants could best align their movement timing when they were performing movements at the same frequency but producing different sounds. Interestingly, their findings suggest that when involved in musical real-time joint actions, the best coordination performance can be achieved when participants segregate self-other auditory information, while integrating self-other visual information. Such findings may inform research in joint action with regards to the complex interplay between inhibiting and merging self- and other- related representations and how this is modulated by the specific sensory modality. Furthermore, the authors discuss theoretical implications with regards to discrete and continuous control processes recruited to support interpersonal coordination during joint action.

One challenging aspect of joint actions is to formulate joint action plans that incorporate both your own and your partners' action possibilities, with their related costs. For individuals involved in joint actions, it becomes necessary at times to evaluate and compare what are the most efficient action alternatives for oneself and the partner. However, individuals in social interactions may be sensible to factors other than instrumental efficiency, for example the fairness of the task distribution among interacting partners. Strachan and Török (2020) (in this special issue) investigate whether individuals in joint actions prioritize action efficiency when it would require partners to contribute unequally to the joint action. Participants involved in a computer task were asked to pass objects to a partner by dragging them on the screen. Crucially, the length of the path and the relative task distribution among co-actors would vary depending on the experimental manipulation. The authors measured participants' preferences and found that, overall, individuals preferred joint action solutions that maximized action efficiency over fairness of task distribution.

### *2.5. Implementation into biological and artificial systems*

The previous section indicates that when we are engaged in a joint action, we need to integrate our partner's actions with our own actions. A key factor for successful joint action is thus joint planning and control. We have already discussed a range of examples of how such control and planning may work at a behavioral level, but how is it implemented at a neural level? Bolt and Loehr (2021) (in this special issue) discuss neuroimaging studies investigating the contribution of the motor system to joint action. By reviewing a range of findings from multiple brain imaging techniques (including fMRI, fNIRS, EEG, TMS), they establish three principles. First, acting in a joint context changes the way the motor system responds to observed action. Second, during joint action, distinct motor representations are associated with self-actions, other-actions, and joint actions. Third, motor processes that serve advanced forms of joint actions reflect the structure of the joint activity, such as the co-actors' roles (e.g., leader versus follower) and whether their actions are coupled in a synchronized or complementary manner in space and/or time.

These principles are of both theoretical and practical importance. An appealing idea is that research in social robotics should build upon these principles (and other principles derived from studying human-human interaction) to improve acceptance and performance in human-robot interactions (Wiese, Metta, & Wykowska, 2017). The logic of the approach is straightforward: if robots are to be treated as social companions, they should evoke the responses and activate the same processes that are typically activated when humans interact with other humans.

Hinz, Ciardo, and Wykowska (2021) (in this issue) adopted this logic in an ERP study of action planning and performance monitoring in human-robot joint action. Participants performed a modified version of the classic Balloon Analogue Risk Task, in which they had to press a key to stop the balloon from inflating and exploding. Performing the task with a robot (joint condition) affected the rise of the Readiness Potential (RP), an electrophysiological marker of early motor preparation. Interestingly, performing the task jointly with the robot also affected outcome monitoring. Specifically, negative outcomes elicited larger feedback-related negativity (FRN) – an ERP component that arise in response to negative outcomes and is modulated by acting in a joint context (e.g., Loehr et al., 2013) – when performing the task with the robot than when performing it alone.

What is more surprising about these results is the ability for humans to engage in joint actions with co-acting robot Cozmo, a small vehicle-shaped commercial robot (Anki Inc., San Francisco). As this robot is not human-shaped, one may postulate that human agents would have difficulty to collaborate with such a robot. The observation that smooth collaboration between Cozmo and human action partners occurs suggests that at least some of the principles of joint action across people extend to interactions with non-humanoid robots.

However, as a study by de la Rosa et al. (2020) (in this issue) shows, it would be wrong to conclude that the visual appearance of the artificial co-actor does not have an influence on joint action planning and control. de la Rosa et al. (2020) demonstrate how the visual appearance of the co-actor, in addition to kinematic-related features, influences motor control. In a novel mixed reality setup, participants executed a high five with a three-dimensional life-size human- or a robot-looking avatar. Movement trajectories and adjustments to perturbations depended on the visual appearance of the avatar (despite the fact that both avatars carried out identical movements), suggesting that control of joint actions is influenced by the human likeness of the human partner.

## 3. General discussion and future directions

This article aimed to accomplish two main goals. First, we hope to have provided an accessible introduction to the field of joint action research. Second, we aimed to situate contributions to our special issue within the broader field of joint action research. We hope that doing so is helpful to new researchers as well as to those already familiar with joint action research.

In the remainder of this article, we will provide some further situating of joint action research as it pertains to some additional theoretical and applied issues. We will also provide some indications of what we think is next in terms of remaining challenges for the field of joint action.

### 3.1. Joint action and embodied cognition

Joint action research forms a powerful example of the tight link between cognition and action. For example, joint actions rely on aligning the perception of other people's actions with our own actions. In this regard, there is clear evidence that the perception of others' actions engages not just our perceptual systems but our motor systems as well. For example, anticipating the reaching and grasping actions of a joint action partner results in similar preparatory activation of the motor system as when one is preparing to perform the same actions oneself. The anticipatory motor activation does not occur when anticipating actions by people who do not take part in the joint action (e.g., Kourtis, Sebanz, & Knoblich, 2010). Employing our own motor systems during action perception is beneficial for joint actions, as it converts observed actions and our own produced actions into a common currency. In this regard, action perception (and consequently joint action) relies on embodied cognitive mechanisms.

Action perception is also involved in higher-level cognitive aspects of joint actions, including the attribution of intentions. Movement kinematics convey a wealth of information about other people's intentions and internal states (C. Becchio, Koul, Ansuini, Bertone, & Cavallo, 2018). By reading such information, we can quickly and often successfully infer and update others' intentions (e.g., Patri et al., 2020). This ability is key to accomplish tasks such as moving furniture lifted by multiple actors as they ascend steep and narrow stairways. Related to this idea, it has been shown that, even while we are not required to interact with another actor, our movement kinematics reveal that we tend to track others' beliefs about for example object locations, even when such beliefs are task-irrelevant (e.g., R.P.R.D. van der Wel, Sebanz, & Knoblich, 2014).

Joint action research does not just speak to notions of embodiment in terms of the link between action perception and activation of our motor systems. In line with another tenet of embodiment, joint action research generally takes a situated approach by considering cognitive activity in a social context (see Wilson, 2002 (Wilson, 2002), for six main tenets of embodiment, and R.P.R.D. van der Wel, Sebanz, and Knoblich (2016) for an application of those tenets to joint action, specifically). Joint action research also tends to take actions as a starting point for understanding cognition, and does so in time-pressured contexts. Joint action research often considers the close interaction between perception, action, and the environment that has been proposed by dynamical systems as well as by radical embodiment approaches. While amenable to embodiment, joint action research has also raised challenges for radical embodiment, however. For example, the evidence for shared task representations and joint memory effects are not easily accounted for based on embodied processing.

In general, then, it is clear that successful joint actions employ cognitive and sensorimotor mechanisms that are shaped by our own action abilities. Future studies on joint action should help to constrain theories of embodiment by clarifying in which sense the different online and offline mechanisms supporting joint action are embodied.

### 3.2. Joint action and social cognition

Joint action research provides a paradigmatic shift away from studying individual brains and minds in isolation. Instead, the field often takes social interaction as a starting point for understanding cognitive and social processes. The approach is, perhaps surprisingly, novel in research on cognitive processes as well as social cognition.

Social cognition includes a range of distinctive processes, including imitation, gaze following, theory of mind, and social learning. These processes are thought to have contributed to the development of language as well. Language forms perhaps the strongest example of why studying minds in socially embedded contexts is crucial, as language specifically exists for interactions with others. The field of joint action research goes beyond the study of verbal interactions by considering other social cognitive processes in socially embedded contexts as well. The contributions to our special issue provide several examples of how this approach translates to a wide range of empirical settings.

One specific example that is not included in our special issue is sensorimotor communication, conceptualized as the process of imposing a communicative intention on a pragmatic action (Pezzulo et al., 2019). Sensorimotor communication operates through subtle changes in movement kinematics. For example, an individual lifting a box may modify their movement kinematics to communicate the information about the weight of the box being lifted (C. Becchio & Panzeri, 2019). In the context of a passing interaction, this information may be used by a receiver to coordinate and apply appropriate grasp forces during object transfer (for example, Mason & MacKenzie, 2005).

### 3.3. Remaining challenges for joint action research

As would be the case for any relatively nascent field of study, many challenges remain in advancing our understanding of joint actions. Rather than providing an extensive overview here, we would like to indicate what we see as some of the most prominent ones.

### 3.4. Scaling up group size

One major challenge concerns the translation of joint action research that has mostly focused on dyadic interactions (i.e., two actors performing some task) to joint actions performed by larger groups. In some way, the question that arises here is analogous to one of the main questions that started joint action research, namely how mechanisms involved in individual cognition scale up to joint actions. Likewise, we should ask how mechanisms involved in dyadic joint actions scale up to group behaviors. Taking this next step is important, as it links action cognition to group dynamics. A question that arises from doing so is when interactions between group members are best characterized by the dynamics of the overall group or by the more local dynamics of a few group members near them. To illustrate this, one could think of a school of fish or a flock of birds as they move in apparent unison. While in those two cases the overall group behavior may be accounted for based on relatively low-level and local interaction dynamics between groups of three or so fish (e.g., Katz, Tunstrom, Ioannou, Huepe, & Couzin, 2011), this is unlikely to be the case for more complex and more goal-directed actions between multiple human agents. In that case, joint action research on larger groups could examine whether co-representation extends beyond dyads, and if so, whether there are particular signature limits to how many people and action components may be co-represented. Other questions arise with respect to the sense of agency for larger scale joint actions as well, as studies in that domain have mostly focused on dyadic interactions as well (e.g., R.P.R.D. van der Wel, 2015). Finally, the interactions between representational and low-level coordination mechanisms rears its head in this context as well.

We would like to highlight two examples of what studies beyond dyadic joint actions may look like, and what they may contribute. One of these examples is a set of studies on multi-agent gaze following. In these studies, F. Capozzi, Cavallo, Furlanetto, and Becchio (2014) had participants watch a scene of a room with blue disks on different locations on the walls. In addition, either one or two avatars stood in the room and gazed towards one of the walls. In case of the two avatars, they looked either at the same position or at two different positions on the wall, implying that they either saw the same or a different number of disks. In each condition, participants needed to verify whether a number that was shown before the scene matched the number of disks visible from different perspectives. In some of the trials, participants were instructed to verify if the number shown matched the number of disks they saw in the room. In other trials, participants were instructed to adopt the perspective of the avatars. In that case, participants verified whether the number shown matched the number of disks the avatar(s) could see. By manipulating the number of avatars and whether the two avatars had the same perspective or not, it was possible to determine whether participants could simultaneously track the visual perspective of multiple agents, and whether this would be as easy to do for two avatars as it would be for one avatar. The results indicated that participants performed better when their own perspective and the perspectives of the avatars implied identical answers. Interestingly, performance significantly decreased when the perspective of the single avatar or the two avatars looking at the same part of the wall implied a different answer. When the two avatars looked at different parts of the wall, the inconsistency did not slow down participants' performance. This finding suggests that people can track the perspective of multiple agents, but they may not spontaneously track multiple viewpoints simultaneously. Follow up experiments (F. Capozzi, Bayliss, & Ristic, 2018) further suggest that as group size increases to five agents, a larger number of agents with a similar perspective to the participant is needed to facilitate performance in this task.

Whether this principle extends to even larger groups still needs to be investigated. Despite these results coming from observational tasks, they carry clear and important implications for the study of representational processes during online joint actions. In fact, one may predict that spontaneous visuo-spatial perspective taking during a joint action might be modulated by the number of potential interactive partners whose perspectives and action possibilities are aligned.

A second example scaled up joint actions to larger groups. In their study, von Zimmermann and Richardson (2016) first asked groups of around 30 participants to read aloud a word list either in unison or while different participants read different parts of the list. Then, they performed a coordination task in which they needed to balance a tightrope walker on a tightrope. This was accomplished by providing each participant with a joystick, for which button presses on each joystick simultaneously influenced the movements of the tightrope walker. The task was to successfully balance the tightrope walker for 30 s. After this task, participants completed a surprise memory test for the words they read, and provided ratings of how affiliated they felt with the group. Results indicated that memory performance as well as affiliation improved from reading the words in synchrony. In addition, results indicated that participants in the synchronous condition responded more readily when the tightrope walker deviated from being upright. The group's responses once the tightrope walker became off-balance were more similar in the synchronous than in the asynchronous group as well. These two studies provide a peek into how one could go about studying mechanisms underlying larger group behavior from a joint action perspective. Much more work is needed to be able to build a proper theory on how features of the individual's actions, the group's actions, the involvement of lower and higher-level mechanisms, and social and cultural aspects of the task context together give rise to large-scale joint performances.

#### *3.4.1. Artificial intelligence and interactive robotics*

Another major challenge for joint action research lies in its possible contribution to the development of interactive artificial agents. While much work has focused on this challenge in the past two decades (see Goswami & Vadakkepat, 2018, for an extensive reference), there are still many fundamental questions that have not been fully addressed. One is whether it is best to implement human-like mechanisms for action production and perception in robots as much as possible, or whether creating artificial agents that interact in a too human-like manner will reduce people's willingness to cooperate with them. In terms of cooperation, questions also arise about how role distributions between the human and artificial agents would be negotiated and implemented in a way that optimizes both the success and the experience of the performed actions.

As an illustrative example of some of the difficulties in developing successful human-robot interaction, there is evidence to suggest that implementing human-like movement kinematics could support successful human-robot interaction. These kinds of kinematics may include smooth velocity profiles that minimize mean square jerk (Flash & Hogan, 1985), as well as early kinematic adjustments such as changes in grasp aperture depending on object size (Jeannerod, 1981) and changes in grasp location based on the action's end goal (e.g., Cohen & Rosenbaum, 2004). From a theoretical perspective, it is reasonable to assume that more human-like robot movement would improve predictability by easing the mapping of perceived kinematics onto a person's own motor system. This mapping could then be used to run simulations of the robot's behavior, in a way that is similar to human-human interaction. In addition, robots could exaggerate certain features of their movements as a form of sensorimotor communication, which has been shown to happen in human-human coordination as well. However, some work also suggests that there is a "sweet spot" in terms of using sensorimotor communication by robotic agents (e.g., Dragan, Lee, & Srinivasa, 2013). When certain kinematic features become too exaggerated, the resulting movements appear difficult to predict. Such lack of predictability in turn seems to damage people's willingness to interact with robotic agents. In general, these kinds of issues indicate that the successful development of joint actions between humans and artificial agents is a complex, multi-facetted challenge.

#### *3.4.2. Developmental and clinical applications*

The potential importance of joint action research in developmental and clinical settings warrants brief mention as well. There is a large literature on how typically developing children develop parallel play and subsequently joint action abilities over the first few years of life. A deeper understanding of typical development can help in identifying irregularities in joint action performance that may provide early markers for developmental delays and disorders. Similarly, irregularities in joint action capabilities may arise in different clinical groups as well. In this regard, it has for example been shown that autistic traits influence interpersonal motor coordination by modulating role-based behavior (Curioni, Minio-Paluello, Sacheli, Candidi, & Aglioti, 2017). Future research could determine whether the study of joint actions may provide insights into clinical manifestations that would not be as apparent from studying patients in isolation.

### 3.5. Conclusion

We would like to leave the reader with an appreciation of the pervasiveness of joint actions. Joint actions form an integral part of the human experience, be it by fostering social connection and providing expressions of culture, or by extending the limits of our own bodies to accomplish goals we could not accomplish alone. The pyramids, the Eiffel Tower, or the grocery store down the street only exist due to the planning and coordination of joint actions across many individuals. By engaging in joint actions, people have made our world what it is today.

## Acknowledgements

The authors would like to thank Natalie Sebanz and Guenther Knoblich for initiating the bi-annual joint action meetings (JAM). These meetings have taken place since 2005. The current special issue is loosely based on the 2019 JAM meeting in Genoa, Italy, with support from the European Union Horizon 2020 Research and Innovation grant agreement no. 824160 (EnTimeMent). Arianna Curioni and Thomas Wolf received support from a grant by the European Research Council (ERC) under the European Union's Seventh Framework Programme [FP7/2007-2013/ERC Grant 609819], project SOMICS.

## References

Becchio, C., Koul, A., Ansuini, C., Bertone, C., & Cavallo, A. (2018, Mar). Seeing mental states: An experimental strategy for measuring the observability of other minds. *Physics of Life Reviews, 24*, 67–80. https://doi.org/10.1016/j.plrev.2017.10.002

Becchio, C., & Panzeri, S. (2019, Mar). Sensorimotor communication at the intersection between kinematic coding and readout comment on "The body talks: Sensorimotor communication and its brain and kinematic signatures" by Giovanni Pezzulo, Francesco Donnarumma, Haris Dindo, Alessandro D'Ausilio, Ivana Konvalinka, Cristiano Castelfranchi. *Physics of Life Review, 28*, 39–42. https://doi.org/10.1016/j.plrev.2019.01.019

Böckler, A., Knoblich, G., & Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1404–1415. https://doi.org/10.1037/a0027523

Bolt, N. K., & Loehr, J. D. (2021). The motor-related brain activity that supports joint action: A review. *Acta Psychologica, 212*, 103218. https://doi.org/10.1016/j.actpsy.2020.103218

Capozzi, F., Bayliss, A. P., & Ristic, J. (2018). Gaze following in multiagent contexts: Evidence for a quorum-like principle. *Psychonomic Bulletin & Review, 25*(6), 2260–2266. https://doi.org/10.3758/s13423-018-1464-3

Capozzi, F., Cavallo, A., Furlanetto, T., & Becchio, C. (2014). Altercentric intrusions from multiple perspectives: Beyond dyads. *PLoS One, 9*(12), Article e114210. https://doi.org/10.1371/journal.pone.0114210

Cohen, R., & Rosenbaum, D. (2004). Where grasps are made reveals how grasps are planned: Generation and recall of motor plans. *Experimental Brain Research, 157*(4), 486–495. https://doi.org/10.1007/s00221-004-1862-9

Constable, M. D., & Knoblich, G. (2020). Sticking together? Re-binding previous other-associated stimuli interferes with self-verification but not partner-verification. *Acta Psychologica, 210*, 103167. https://doi.org/10.1016/j.actpsy.2020.103167

Cross, L., Michael, J., Wilsdon, L., Henson, A., & Atherton, G. (2020). Still want to help? Interpersonal coordination's effects on helping behaviour after a 24 hour delay. *Acta Psychologica, 206*, 103062. https://doi.org/10.1016/j.actpsy.2020.103062

Curioni, A., Minio-Paluello, I., Sacheli, L. M., Candidi, M., & Aglioti, S. M. (2017). Autistic traits affect interpersonal motor coordination by modulating strategic use of role-based behavior. *Molecular Autism, 8*(1), 23. https://doi.org/10.1186/s13229-017-0141-0

de la Rosa, S., Meilinger, T., Streuber, S., Saulton, A., Fademrecht, L., Quiros-Ramirez, M. A., … Cañal-Bruland, R. (2020). Visual appearance modulates motor control in social interactions. *Acta Psychologica, 210*, 103168. https://doi.org/10.1016/j.actpsy.2020.103168

Dragan, A. D., Lee, K. C. T., & Srinivasa, S. S. (2013). *Legibility and predictability of robot motion proceedings of the 8th ACM/IEEE international conference on human-robot interaction, Tokyo, Japan, 301–308*.

Eisenberg, N. (1982). *The development of prosocial behavior*. New York, NY: Academic Press.

Eskenazi, T., Doerrfeld, A., Logan, G. D., Knoblich, G., & Sebanz, N. (2018). Your words are my words: Effects of acting together on encoding. *Quarterly Journal of Experimental Psychology, 66*(5), 1026–1034. https://doi.org/10.1080/17470218.2012.725058

Flash, T., & Hogan, N. (1985). The coordination of arm movements: An experimentally confirmed mathematical model. *The Journal of Neuroscience, 5*(7), 1688–1703. https://doi.org/10.1523/jneurosci.05-07-01688.1985

Gobel, M. S., & Giesbrecht, B. (2020). Social information rapidly prioritizes overt but not covert attention in a joint spatial cueing task. *Acta Psychologica, 211*, 103188. https://doi.org/10.1016/j.actpsy.2020.103188

Goswami, A., & Vadakkepat, P. (2018). *Humanoid*. Robotics: *A Reference*. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7194-9

Gvirts Problovski, H. Z., Lavi, D., Yozevitch, R., Sherman, M., Hagay, Y., & Dahan, A. (2021). Impairments of interpersonal synchrony evident in attention deficit hyperactivity disorder (ADHD). *Acta Psychologica, 212*, 103210. https://doi.org/10.1016/j.actpsy.2020.103210

Hadley, L. V., Fisher, N. K., & Pickering, M. J. (2020). Listeners are better at predicting speakers similar to themselves. *Acta Psychologica, 208*, 103094. https://doi.org/10.1016/j.actpsy.2020.103094

Haugeland, J. (1991). Representational genera. In W. Ramsey, S. P. Stich, & D. E. Rumelhart (Eds.), *Developments in connectionist theory. Philosophy and connectionist theory (p. 61–89)*. Lawrence Erlbaum Associates, Inc.

Hinz, N.-A., Ciardo, F., & Wykowska, A. (2021). ERP markers of action planning and outcome monitoring in human – Robot interaction. *Acta Psychologica, 212*, 103216. https://doi.org/10.1016/j.actpsy.2020.103216

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2002). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*(5), 849–878. https://doi.org/10.1017/s0140525x01000103

Jeannerod, M. (1981). Intersegmental coordination during reaching at natural visual objects. In J. Long, & A. Baddeley (Eds.), *Attention and performance IX (pp. 153–168)*. Lawrence Erlbaum.

Kato, M., Kitagawa, N., Kimura, T., Takano, Y., Takagi, T., Hirose, H., & Kashino, M. (2020). Walking and talking independently predict interpersonal impressions. *Acta Psychologica, 210*, 103172. https://doi.org/10.1016/j.actpsy.2020.103172

Katz, Y., Tunstrom, K., Ioannou, C. C., Huepe, C., & Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences, 108*(46), 18720–18725. https://doi.org/10.1073/pnas.1107583108

Kourtis, D., Sebanz, N., & Knoblich, G. (2010). Favouritism in the motor system: Social interaction modulates action simulation. *Biology Letters, 6*(6), 758–761. https://doi.org/10.1098/rsbl.2010.0478

Kuhn, M. H., & McPartland, T. S. (1954). An empirical investigation of self-attitudes. *American Sociological Review, 19*(1), 68–76. https://doi.org/10.2307/2088175

Lee, H., Launay, J., & Stewart, L. (2020). Signals through music and dance: Perceived social bonds and formidability on collective movement. *Acta Psychologica, 208*, 103093. https://doi.org/10.1016/j.actpsy.2020.103093

Lelonkiewicz, J. R., & Gambi, C. (2020). Making oneself predictable in linguistic interactions. *Acta Psychologica, 209*, 103125. https://doi.org/10.1016/j.actpsy.2020.103125

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience, 25*(7), 1049–1061. https://doi.org/10.1162/jocn_a_00388

Mason, A. H., & MacKenzie, C. L. (2005). Grip forces when passing an object to a partner. *Experimental Brain Research, 163*(2), 173–187. https://doi.org/10.1007/s00221-004-2157-x

Michael, J., McEllin, L., & Felber, A. (2020). Prosocial effects of coordination – What, how and why? *Acta Psychologica, 207*, 103083. https://doi.org/10.1016/j.actpsy.2020.103083

Nafcha, O., Morshed-Sakran, A., Shamay-Tsoory, S., & Gabay, S. (2020). The effect of co-actor group membership on the social inhibition of return effect. *Acta Psychologica, 208*, 103119. https://doi.org/10.1016/j.actpsy.2020.103119

Néda, Z., Ravasz, E., Vicsek, T., Brechet, Y., & Barabási, A. L. (2000). Physics of the rhythmic applause. *Physical Review E, 61*(6), 6987–6992. https://doi.org/10.1103/PhysRevE.61.6987

Patri, J.-F., Cavallo, A., Pullar, K., Soriano, M., Valente, M., Koul, A., … Becchio, C. (2020). Transient disruption of the inferior parietal lobule impairs the ability to attribute intention to action. *Current Biology, 30*, 4594–4605. https://doi.org/10.1016/j.cub.2020.08.104

Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., & Castelfranchi, C. (2019). The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of Life Reviews, 28*, 1–21. https://doi.org/10.1016/j.plrev.2018.06.014

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann, & W. Prinz (Eds.), *Relationships between perception and action (pp. 167–201)*. Springer. https://doi.org/10.1007/978-3-642-75348-0_7.

Prinz, W. (2015). Task representation in individual and joint settings. *Frontiers in Human Neuroscience, 9*, 268. https://doi.org/10.3389/fnhum.2015.00268

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science, 26*(6), 867–891. https://doi.org/10.1016/j.humov.2007.07.002

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human Movement Science, 30*(5), 834–845. https://doi.org/10.1016/j.humov.2010.05.014

Schmitz, L., Knoblich, G., Deroy, O., & Vesper, C. (2021). Crossmodal correspondences as common ground for joint action. *Acta Psychologica, 212*, 103222. https://doi.org/10.1016/j.actpsy.2020.103222

Scott Jordan, J., Schloesser, D. S., Bai, J., & Abney, D. (2018). Multi-scale contingencies during individual and joint action. *Topics in Cognitive Science, 10*(1), 36–54. https://doi.org/10.1111/tops.12304

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences, 10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

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. https://doi.org/10.1037/0096-1523.31.6.1234

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 326–332. https://doi.org/10.1037/0096-1523.29.2.326

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. P. T. G. Reeve (Ed.), *Advances in psychology. Stimulus-response compatibility: An integrated perspective (Vol. 65, pp. 31–86)*. North-Holland.

Simon, J. R., Hinrichs, J. V., & Craft, J. L. (1970). Auditory S-R compatibility: Reaction time as a function of ear-hand correspondence and ear-response-location correspondence. *Journal of Experimental Psychology, 86*(1), 97–102. https://doi.org/10.1037/h0029783

Strachan, J. W. A., & Török, G. (2020). Efficiency is prioritised over fairness when distributing joint actions. *Acta Psychologica, 210*, 103158. https://doi.org/10.1016/j.actpsy.2020.103158

van der Wel, R. P. R. D. (2015). Me and we: Metacognition and performance evaluation of joint actions. *Cognition, 140*, 49–59. https://doi.org/10.1016/j.cognition.2015.03.011

van der Wel, R. P. R. D., & Fu, E. (2015). Entrainment and task co-representation effects for discrete and continuous action sequences. *Psychonomic Bulletin & Review, 22*(6), 1685–1691. https://doi.org/10.3758/s13423-015-0831-6

van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others' beliefs? Evidence from a continuous measure. *Cognition, 130*(1), 128–133. https://doi.org/10.1016/j.cognition.2013.10.004

van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2016). A joint action perspective on embodiment. In M. H. F. Y. Coello (Ed.), *Foundations of embodied cognition: Conceptual and interactive embodiment (pp. 165–181)* (Routledge/Taylor & Francis Group).

von Zimmermann, J., & Richardson, D. C. (2016). Verbal synchrony and action dynamics in large groups. *Frontiers in Psychology, 7*, 2034. https://doi.org/10.3389/fpsyg.2016.02034

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? *Neuroscience Letters, 385*(2), 99–104. https://doi.org/10.1016/j.neulet.2005.05.013

Welsh, T. N., Reid, C., Manson, G., Constable, M. D., & Tremblay, L. (2020). Susceptibility to the fusion illusion is modulated during both action execution and action observation. *Acta Psychologica, 204*, 103028. https://doi.org/10.1016/j.actpsy.2020.103028

Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: Using neuroscientific methods to make robots appear more social. *Frontiers in Psychology, 8*, 1663. https://doi.org/10.3389/fpsyg.2017.01663

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636. https://doi.org/10.3758/bf03196322

Robrecht P.R.D. van der Wel[a,*], Cristina Becchio[b], Arianna Curioni[c], Thomas Wolf[c]

[a] *Department of Psychology, Rutgers University Camden, NJ, USA*

[b] *C'MoN, Cognition, Motion and Neuroscience Unit, Fondazione Istituto Italiano di Tecnologia, Genova, Italy*

[c] *Department of Cognitive Science, Central European University, Vienna, Austria*

[*] Corresponding author.

*E-mail address:* r.vanderwel@rutgers.edu (R.P.R.D. van der Wel).