

# Taking apart what brings us together: The role of action prediction, perspective-taking, and theory of mind in joint action

Quarterly Journal of Experimental Psychology  
1–16

© Experimental Psychology Society 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/17470218211050198

qjep.sagepub.com



Lucia Maria Sacheli<sup>1</sup> , Elisa Arcangeli<sup>2</sup>,  
Desiré Carioti<sup>2</sup>, Steve Butterfill<sup>3</sup>  
and Manuela Berlingeri<sup>2,4</sup>

## Abstract

The ability to act together with others to achieve common goals is crucial in life, yet there is no full consensus on the underlying cognitive skills. While influential theoretical accounts suggest that interaction requires sophisticated insights into others' minds, alternative views propose that high-level social skills might not be necessary because interactions are grounded on sensorimotor predictive mechanisms. At present, empirical evidence is insufficient to decide between the two. This study addressed this issue and explored the association between performance at joint action tasks and cognitive abilities in three domains—action prediction, perspective-taking, and theory of mind—in healthy adults ( $N=58$ ). We found that, while perspective-taking played a role in reading the behaviour of others independently of the social context, action prediction abilities specifically influenced the agents' performance in an interactive task but not in a control (social but non-interactive) task. In our study, performance at a theory of mind test did not play any role, as confirmed by Bayesian analyses. The results suggest that, in adults, sensorimotor predictive mechanisms might play a significant and specific role in supporting interpersonal coordination during motor interactions. We discuss the implications of our findings for the contrasting theoretical views described earlier and propose a way they might be partly reconciled.

## Keywords

Joint action; action prediction; perspective-taking; theory of mind; social cognition

Received: 4 June 2021; revised: 3 August 2021; accepted: 1 September 2021

## Introduction

Whatever we do, interaction with others can rarely be avoided. In passing a ticket for inspection, receiving a coffee, mutually turning your bodies to pass each other in a corridor, and in myriad other situations, you are engaged in what could be defined as “small-scale social interactions” (Sinigaglia & Butterfill, 2020); the term indicates an interaction which is fleeting, would not typically involve advance deliberation, and is related as a means to some larger end, such as travelling to Milan, preparing to give a talk, or walking out of the theatre. Given the ubiquitous presence of small-scale interactions in everyday life, the quest to understand them spans several disciplines (including philosophy, developmental psychology, and cognitive science) and involves multiple approaches. However, the

<sup>1</sup>Department of Psychology and Milan Center for Neuroscience (NeuroMi), University of Milano-Bicocca, Milan, Italy

<sup>2</sup>Department of Humanistic Studies (DISTUM), University of Urbino Carlo Bo, Urbino, Italy

<sup>3</sup>Department of Philosophy, University of Warwick, Coventry, UK

<sup>4</sup>Center of Developmental Neuropsychology, ASUR Marche, Pesaro, Italy

### Corresponding author:

Lucia Maria Sacheli, Department of Psychology and Milan Center for Neuroscience (NeuroMi), University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy.  
Email: lucia.sacheli@unimib.it

Manuela Berlingeri, Department of Humanistic Studies (DISTUM), University of Urbino Carlo Bo, Via Bramante, 17, 61029 Urbino, Italy.  
Email: manuela.berlingeri@uniurb.it

findings emerging from these disciplines seem, as yet, difficult to reconcile.

It has long been established that navigating the social world requires social cognition, defined as the whole set of cognitive capabilities that allow us to put ourselves “in other people’s shoes” by making sense of their behaviour and modifying our behaviour accordingly (Adolph, 2001; Frith & Frith, 2003). Core cognitive components of social cognition are perspective-taking and theory of mind, i.e., the ability to mentally represent how another person perceives the space at a specific time (visuospatial perspective-taking, Newcombe, 1989) and to associate others’ behaviours with the mental states that have generated them (theory of mind, Baron-Cohen et al., 2001). Exercising these social skills involves inferring, based on the contextual information, how other people perceive the world, or what they think. Leading philosophical accounts (Bratman, 1992, 1999, 2014; Carpenter, 2009) postulate that *inferential social skills* are also at the basis of successful interactions, as interactions require *shared intentions* and the agents’ ability to represent the partner’s mental states and structure bargaining with him or her based on such representations.

Potentially supporting these ideas, there is evidence that social cognition plays a significant role in modulating the participants’ motor behaviours in what might be interpreted as motor interactions (Brass et al., 2009; Sowden & Catmur, 2015; Spengler et al., 2009). These studies do hint that social cognition may be important in small-scales interactions, although they required agents to act while observing another person’s (irrelevant) action, thus making the results interpretation more difficult.

Indeed, when people engage in small-scale interactions, they do not merely observe, interpret, and react to others; instead, *they act together* in joint actions and coordinate to achieve common goals (Sebanz et al., 2006). Thus, one might suggest that on-line social exchanges might also (or mainly) involve sensorimotor processes. Indeed, interactive gestures are engaging (Curioni et al., 2020; Sartori et al., 2009) and seem to call for complementary responses without requiring complex inferential computations, as when passing or receiving an object. Moreover, developmental psychologists suggest that children show the first signs of structured cooperative skills in the first years of life (Brownell, 2011; Brownell et al., 2006; Meyer et al., 2010; Warneken et al., 2012; Warneken & Tomasello, 2007), when their grasp of other people’s intentions and mental life is likely far from being fully developed. In these early interactions, sensorimotor predictive mechanisms seem to play a major role, in particular, by underpinning the ability to anticipate others’ goals while observing their movements (Meyer et al., 2015).

This empirical evidence from developmental research is in line with the suggestion that interaction requires the predictive ability to anticipate the partner’s behaviour and

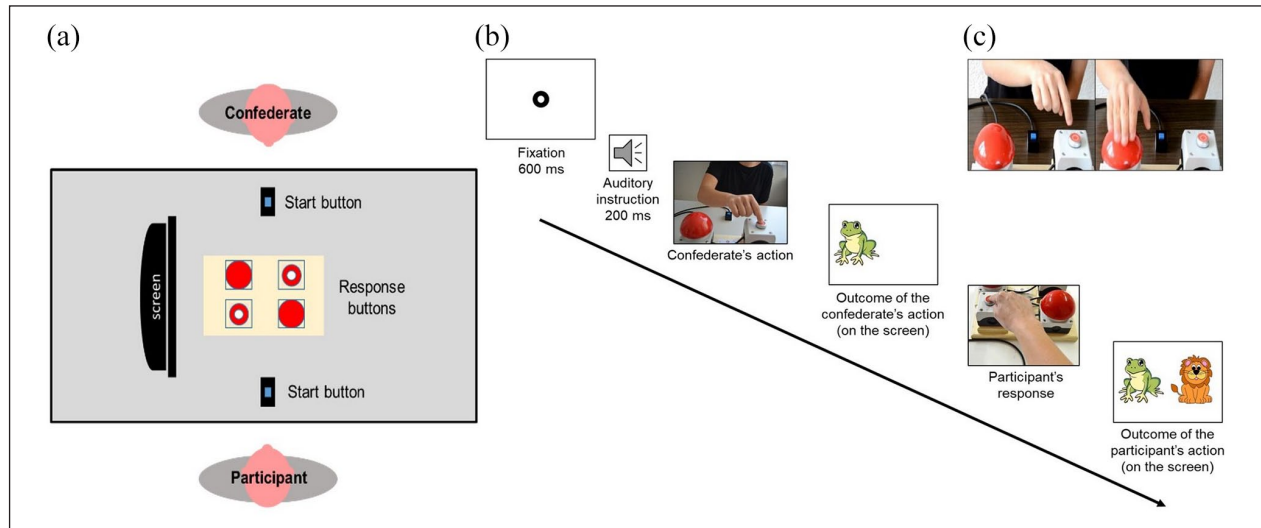
adapt accordingly (Bekkering et al., 2009; Knoblich & Jordan, 2003; Vesper et al., 2013) to facilitate the achievement of a shared goal (Butterfill, 2012). Although, more broadly, inferences and predictions regarding others’ gesture may depend on the bidirectional flow of information between brain networks processing semantic or contextual and sensorimotor information (Finisguerra et al., 2020; Kilner, 2011), research on motor interaction focusses on the latter and explores to what extent the (sensorimotor) predictive and monitoring mechanisms that govern individual motor control play a role in interpersonal coordination (see Candidi et al., 2015; Pesquita et al., 2018; Pezzulo et al., 2017) when both the agent’s and the partner’s actions are represented within an integrated, dyadic motor plan (Sacheli, Arcangeli, & Paulesu, 2018). The involvement of such sensorimotor predictive mechanisms may explain how reciprocal motor adjustments can quickly occur between interacting partners. For instance, when moving a table together from one room to another one, we can broadly predict our partner’s next step (e.g., what direction he or she will take) based on the knowledge of the overarching goal (e.g., going to the kitchen) and direct our actions towards the same goal; we can also apply motor computations to obtain fine-grained predictions, e.g., based on the partner’s kinematics, and adjust our movements accordingly with high degrees of temporal precision (Pesquita et al., 2018; Pezzulo et al., 2017). Thus, a dyadic motor plan structures interpersonal coordination by channelling (motor) predictions on what the partner will do based on a representation of the shared goal (Sacheli, Arcangeli, & Paulesu, 2018) and possibly without requiring inferential processes but instead involving sensorimotor predictive processes only.

Thus, there might be at least two broad classes of processes that sustain motor interactions; interactions might be based on either the *inferential* ability to represent shared intentions, based on theory of mind (Bratman, 1992, 1999, 2014; Carpenter, 2009), or the *sensorimotor* capacity to apply sensorimotor predictions triggered by the representation of shared goals (Butterfill, 2012; Pesquita et al., 2018; Sacheli, Arcangeli, & Paulesu, 2018). These different stand-points are represented in Table 1. To date, the relative role of inferential and sensorimotor predictive mechanisms in joint action is far from clear. One way to approach this issue is testing whether inter-individual differences in the relevant cognitive abilities (i.e., either predict others’ action goals or interpret their perspective and mental states) have an impact on performance in joint action (JA) tasks, something that we aimed to do here.

This study aims at filling this gap and test, in a sufficiently large sample of healthy adults, whether the individual sensorimotor skills (action prediction) and inferential skills (theory of mind) might moderate participants’ performance in a simple JA task. We also added perspective-taking as a possible moderator because it is a

**Table 1.** The table illustrates what would be the theoretical implications of JA being based on shared intentions versus shared goals. While JA based on shared intentions would imply inferential skills like theory of mind, JA based on shared goals set their roots in sensorimotor processes like action prediction.

Interactions based on	Definitions	Cognitive processes
Shared intentions (Bratman, 1992, 1999, 2014)	A structure of mental states involving knowledge, on the part of each agent involved in the interaction that they all know that they each intend to do something together.	Inferential processes (theory of mind)
Shared goals (Butterfill, 2012)	An outcome in which the agents involved in a joint action coordinate their actions to achieve together.	Sensorimotor processes (action prediction)



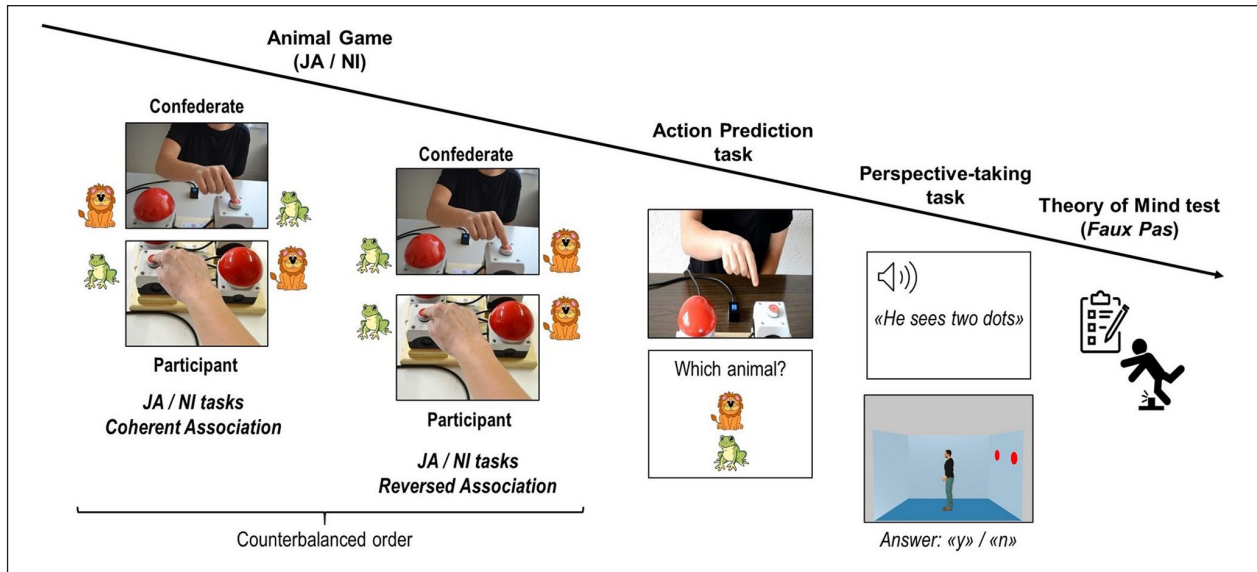
**Figure 1.** The experimental set-up and trial-timeline of the Animal Game which were identical in the Joint Action (JA) and Non-Interactive (NI) tasks. (a) The participant and the experimenter's confederate were seated one in front of the other. (b) The task required the participants to “make the animal meet on the screen” (JA task) or “make a specific animal appear on the screen yet respecting the turn-taking nature of the task” (NI task); the confederate always acted first and then it was the participant's turn. (c) The shape of the two buttons afforded two different actions to operate on: a whole-hand press on the bigger button and an index-finger press on the smaller button.

human-specific social ability linked to theory of mind (Tomasello et al., 2005) that is yet also (or mainly) based on perceptual processes (Ward et al., 2019; but see Pavlidou et al., 2019). Implicit visual perspective-taking and verbal theory of mind abilities can be seen as two dissociable components along the continuum of the cognitive abilities that constitute social cognition.

We tested 58 young, healthy participants in a simple turn-taking motor task adapted from the one described by Sacheli, Meyer, et al. (2019). The task required the participants to play together with a partner (the experimenter's confederate) to lead animal-like cartoon characters “meet” on the screen by pressing one of two different buttons (Figure 1). The task could be played in an Interactive (JA) and a Non-Interactive (NI) version; both required the same turn-taking button-press movements, but the partner's behaviour was irrelevant in the NI task, while it was relevant in the JA task because the two partners had to coordinate. Moreover, both the JA and NI tasks required the

participants to play with a partner whose buttons functioned either in the same (Coherent Action-Outcome Association condition) or in the opposite way than the participant's ones (Reversed Action-Outcome Association condition). Namely, in both tasks, there was a condition in which the button-animal association was coherent between the partner and the participant, and one in which it was reversed.

Although a simple motor task, interaction with a partner in the JA task might, unlike the NI task, require taking into account the partner's perspective, leading to a behavioural cost when interacting with a partner having a reversed button-animal association. Accordingly, previous studies applying similar experimental designs showed an interaction effect between the Task (JA/NI) and the Action-Outcome Association (Coherent/Reversed), showing decay in the participants' performance in the Reversed Action-Outcome Association condition that was selective for the JA task (Sacheli, Arcangeli, & Paulesu, 2018;



**Figure 2.** The figure illustrates the structure of the Experiment. First, the participants performed the Animal Game: they played the JA and NI tasks, in counterbalanced order, while having a button-animal (Action-Outcome) association that was either Coherent or Reversed between the confederate and the participant (counterbalanced order between the participants). Then, they performed the Prediction task, the Implicit Perspective-Taking task, and the Faux Pas test measuring theory of mind. In the analyses, we first analysed each task separately and then extracted three indexes from (1) the Prediction task, (2) the Implicit Perspective-Taking task, and (3) the Faux Pas test measuring theory of mind; these indexes were entered as moderators in the analysis of the performance achieved at the Animal Game. See the “Methods” section for details.

Sacheli et al., 2021; Sacheli, Meyer, et al., 2019). In this study, we aimed to replicate such a JA-specific effect and explore its possible predictors. We aimed to examine whether the selective decay of performance in Reversed Association trials in the JA task could be moderated by either sensorimotor (action prediction) or inferential skills (visual perspective-taking or theory of mind). After completion of the task, we, therefore, measured in all participants (1) sensorimotor (action prediction) skills, (2) perspective-taking skills, and (3) theory of mind abilities.

See Figure 2 for a schematic representation of the structure of the experiment.

## Methods

### Participants

Fifty-eight healthy adult participants took part in the study (37 female, age range 18–35, mean age  $22.31 \pm 3.12$ ). All participants were right-handed (self-report), reported normal or corrected to normal vision, and were naive about the purposes of the experiment. They gave their written informed consent to take part in the study and were debriefed on the purposes of the experiment at the end of the experimental procedure. The experimental protocol was approved by the Ethics Committee of the University of Urbino “Carlo Bo.” The participants provided their written informed consent, in accordance with the ethical standards of the 1964 Declaration of Helsinki and later amendments.

Sample-size was determined by a power analysis implemented in the software jamovi (version 1.1.8.0, <https://www.jamovi.org>, The jamovi project, 2020) and based on the data of a previous study (Sacheli, Meyer, et al., 2019) that used a joint action task similar to the one employed here. The results of a paired-sample *t*-test performed on these data revealed a behavioural modulation of individuals’ performance depending on the feature of the interactive task with a  $d=0.6$ . The power analysis showed that, with  $\alpha=.01$  and statistical power at  $1-\beta=.95$ , we needed a sample size of  $N=53$  to be sufficiently sure to replicate such an effect. This sample-size was also suitable for our correlational approach. We recruited participants with public announcement in university courses; 58 students responded and were thus tested in the final sample.

### Structure of the study

The study included three tasks and one questionnaire, which were administered in fixed order between participants. First, the participants performed the Animal Game, which included two different versions of a JA and an NI tasks performed in different sessions and administered in counterbalanced order between participants. The two versions differed depending on the functioning of the partner’s buttons during the task (see below). Participants started with one version, performed both the JA and NI tasks (in counterbalanced order between participants), and then performed the second version of both tasks in a



second session. Afterwards, each participant performed the Prediction Task (see below), the Implicit Perspective-Taking task (revised from Samson et al., 2010; Surtees & Apperly, 2012), and the Faux Pas (FP) test (Stone et al., 1998, 2018; Italian translation by Massaro, Liverta-Sempio, & Marchetti). The order of the tasks was maintained fixed in all participants. Details on each task are reported below.

In the three computer-based tasks (Animal Game, the Prediction task, and the Implicit Perspective-Taking task), stimuli presentation and randomisation were controlled by Opensesame software (Mathôt et al., 2012), and participants were required to respond as fast and accurately as possible. The FP test was presented in a paper-and-pencil version.

In this section, we first describe each task separately (the Animal Game, the Prediction task, and the Implicit Perspective-Taking task, as well as the FP test) and how each was analysed to check that the expected results (based on previous studies) were individually replicated. Then, we describe the multiple regression analyses that we performed to explore whether the performance at the Animal Game could be moderated by either sensorimotor (action prediction) or inferential skills (visual perspective-taking or theory of mind) as measured by the other tasks.

### Animal game

**Stimuli and apparatus.** The Animal Game was performed with a set-up similar to the one described in the paper by Sacheli and colleagues (2019). Participants were seated in front of a rectangular table ( $\approx 80 \times 120$  cm). The experimenter's confederate, who played as interaction partner, was seated at the opposite side of the table, in front of him or her. A custom-made response box including two pairs of buttons (BrainTrends, Ltd, <https://www.braintrends.it/>) was located on the table, one pair in front of the confederate and one in front of the participant (Figure 1a). The two buttons forming a pair had the same height (10 cm) but different dimensions (8 and 2 cm diameter) and thus, afforded different responses to be operated on; the bigger button could be pressed only with a whole-hand press and the smaller button could be pressed only with a single-finger press (Figure 1). A 15.6-inch laptop screen was located in the middle of the table to the left of the buttons from the participant's perspective, tilted  $\approx 100^\circ$  and oriented in such a way that both the participant and the confederate could optimally see the visual stimuli appearing on it. The buttons were placed at the centre of the table, at  $\approx 50$  cm from the confederate and the participant. In half of the sample ( $n = 30$ ), the configuration of buttons was identical for the participants and the confederate, i.e., they both had their bigger button at their right. This implied the confederate's bigger button being placed on the participant's left side. In the other half of the sample ( $n = 28$ ), the

buttons were placed in a mirror configuration, so that both the participant's and the confederate's bigger buttons were on the participant's left side.

Both the confederate and the participant had to keep their right index-finger on a starting-button ( $1.3 \times 1.3$  cm) placed at the edge of the table and  $\approx 5$  cm on their right. Visual stimuli consisted of images of an animal designed in a cartoon-like fashion (frog or lion, Figure 1b). Both the participant and the confederate received auditory instructions through headphones. Auditory instructions consisted of the following four different sounds: a frog cry, a lion cry, and the Italian words corresponding to "same" (it. "stesso") and "different" (it. "diverso"). All sounds had a duration of 500 ms and comparable intensity.

**Procedure.** The Animal Game included two tasks, JA and NI tasks.

**JA task.** In the JA task, the auditory instructions provided at the beginning of each trial required the participants to press the button which would make the "same" or "different" animal appear on the screen. The order of same or different trials alternated in the different eight-trial blocks and was counterbalanced between the participants. At each trial, both the confederate and the participant had to keep their right hand on the starting-button. The trial started with a fixation cross. After a variable delay (500–900 ms), the confederate heard the auditory instruction indicating which animal (frog or lion) she had to produce on the screen; concomitantly, the auditory instruction was also delivered to the participant and consisted of the word "same" (it. "stesso") or "different" (it. "diverso"). After the auditory instruction, the confederate performed her action; the animal appearing on the screen following the confederate's action was the go-signal for the participant, who could then complete the action by making the second animal appear on the screen (Figure 1b). If the participant's response was correct, the correct animal appeared on the screen, otherwise, a black dot appeared on the screen to provide a negative feedback. For instance, a trial could start with the following instructions: a lion cry delivered to the confederate and the word "same" to the participant. Here, the confederate would first perform a whole-hand press on the bigger button to lead the lion appear on the screen; as soon as the lion appeared, the participant would then perform a whole-hand press on the bigger button to lead a second lion appear on the screen (as mentioned earlier, for the participant, the bigger button was always associated with the lion and the smaller button with the frog).

**NI task.** The NI task was identical to the JA task in its perceptual features and motor requirements. The only difference consisted of the participant's auditory instructions; here, the participants heard, concomitantly to the confederate's instruction, an animal cry indicating which animal

(frog or lion) the participant had to produce on the screen, independently of the animal produced by the confederate. The confederate's action was yet the go-signal for the participants, so that they were anyway required to follow the turn-taking format of the task, as in the JA one. Unbeknownst to participants, in different blocks, the instructions lead them to make the same or different animal appear on the screen 50% of the times, similarly to what happened in the JA task. As in the JA task, if the participant's response was correct, the correct animal appeared on the screen, otherwise, a black dot appeared on the screen instead, to provide a negative feedback. For instance, a trial could start with the following instructions: a lion cry delivered to both the confederate and the participant. Here, the confederate would first perform a whole-hand press on the bigger button to lead the lion appear on the screen; as soon as the lion appeared, the participant would then perform a whole-hand press on the bigger button to lead a second lion appear on the screen. Thus, although instructions are different as compared with the JA task, the motor requirements and perceptive feedbacks are identical in the two tasks.

*Manipulation of button-animal (action-outcome) association.* Both the JA and NI tasks were presented in two versions, which differed depending on the confederate's button-animal association. For the participant, a whole-hand press on the big button produced a lion of the screen, and a single-finger press on the small-button produced a frog on the screen; this association was kept constant throughout the task. Instead, the button-animal (i.e., Action-Outcome) association could change. That is, while in the Coherent Action-Outcome Association (CoA) version, the confederate's association between each button (big or small) and animal (frog or lion) was identical to the participant's one, in the Reversed Action-Outcome Association (RevA) version, such an association was reversed in the confederate (see Figure 2). As a result, we obtained a 2 (Task: JA vs. NI) by 2 (Association: CoA vs. RevA) factorial design.

All participants performed one of the two versions of the two tasks first, and then the other version (e.g., a possible task order could be JA-CoA, NI-CoA, JA-RevA, NI-RevA). Before starting each version, the specific confederate's button functioning was explicitly described to the participant. To highlight the difference between the versions, a yellow sheet was placed in front of the buttons during the Reversed Association sessions. Importantly, the reversal of the button-animal association implied that, in the JA task, the participants had to bear in mind the specific partner's button functioning if they wanted to anticipate what animal would appear on the screen by observing the partner's movements. This might require both inferential skills (i.e., putting oneself "in the confederate's shoes") and predictive abilities (to predict the incoming animal from observing the confederate's action).

Each experimental condition (JA-CoA, NI-CoA, JA-RevA, NI-RevA) included 32 trials, divided in four eight-trial blocks requiring participants to press 50% of the time the big or small button in randomised order by following task-specific instructions.

Each task started with an eight-trial training to familiarise with the set-up and instructions. In both tasks, the confederate always acted first before it was the participant's turn. The partner's role was randomly played by the one of two possible female confederates.

### Prediction task

*Stimuli.* The stimuli consisted of pictures showing the confederate's hand part way through the act of pressing one of the two buttons (implied-motion image, see Figure 2). The picture could show the hand after 1/3 or 2/5 (Difficult condition) and 3/5 or 2/3 (Easy condition) of the movement time. We selected implied-motion images to elicit sensorimotor predictive processes based on previous neurophysiological studies showing that these stimuli are able to elicit neural activity in the sensorimotor system (Avenanti et al., 2013; Urgesi et al., 2006, 2010).

*Procedure.* Each trial started with a fixation dot displayed on the screen for 500 ms, followed by a start-position image that lasted 1 s and by an implied-motion image shown for 200 ms. Finally, the response image was shown, and the participants had to respond indicating which animal (frog or lion) would appear on the screen as a consequence of the agent's action by pressing the "b" key to respond "lion" and the "y" key to respond "frog" (the key-animal association was counterbalanced between the participants). The time-delay between the instant at which the response image was shown on the screen and the participants' button-press was measured as the participant's response time (RT). The task included 24 Easy and 24 Difficult trials, presented in randomised order. After 24 trials, the participants were allowed to have a short break. Participants performed four practice trials before starting the task. The experiment was presented on the same laptop computer used for the Animal Game.

### Implicit perspective-taking task

*Stimuli.* The task was a shorter version of the one first proposed by Samson and colleagues (2010), and similar to the one described by Surtees and Apperly (2012). This task has been proposed as an instrument to investigate Level 1 visual perspective-taking, that is, the human ability to spontaneously process how others perceive the world. Some controversy arose on whether the results obtained from such tasks indeed depend on perspective-taking or more domain general cognitive processes like memory and attention ("submentalising" account by Heyes, 2014;

Santiesteban et al., 2014). However, we employed the task based on recent evidence of a genuine role of perspective-taking in it (Furlanetto et al., 2016), even when controlling for attentional factors, at least as far as reaction times are used as dependent measure (Holland et al., 2021).

Stimuli consisted of pictures of a cartoon avatar standing in a cartoon room with dots on the wall (see Figure 2). Male participants watched a male avatar and female participants a female avatar.

**Procedure.** As described in the original papers (Samson et al., 2010; Surtees & Apperly, 2012), the task included a full-factorial design with four experimental conditions (made by the combination of Self-/Other- and Consistent/Inconsistent conditions, see below). On *Self*-trials, participants judged the number of dots they could see on the walls of the picture. On *Other*-trials, participants judged how many dots could be seen by the cartoon avatar in the picture. On *Consistent* trials the avatar could see the same number of dots as the participant. On *Inconsistent* trials the avatar's position in the room meant that she or he saw fewer dots.

On each trial, the participants viewed two subsequent fixation stimuli (a smiling face [600ms] and a fixation point [600ms]) followed by a 1.8s auditory stimulus (either "He/She sees N" or "You see N," where *N* ranged from 1 to 3), and then the test picture depicting an avatar in a room with 1 to 3 dots on the walls. Participants pressed one of two keys (y or b) to indicate whether the auditory stimulus correctly described the picture (y) or not (b). The sentence stimuli were matched across Self ("You see" sentence) and Other ("He/She sees" sentence) trials. Independently of consistency (i.e., both in Consistent and Inconsistent trials), 50% of the trials required a "yes" response (because the auditory stimulus matched the picture) and the other 50% required a "no" response (because the auditory stimulus did not match the picture).

The participants performed 96 test trials, divided in 24-trial blocks; in half of the trials, Self and Other perspectives were consistent (50% Self/Other), and in the remaining half, they were inconsistent (50% Self/Other). Self- and Other-trials were pseudo-randomly mixed within each 24-trial block so that no block contained more than three trials in a row without a change in consistency, perspective, and response button. The experiment was presented on the same laptop computer used for the Animal Game.

### Faux Pas test

**Stimuli and procedure.** The FP test (Stone et al., 1998) requires the recognition of FP in 20 stories, i.e., situations in which someone mistakenly says something they should not have. Only 10 out of the 20 stories actually contain a FP, while the other 10 are control stories. FP can consist of intentional and non-intentional actions from multiple

perspectives, and their identification makes it possible to determine the appreciation of false beliefs, emotional states, and intentions, while the presence of control stories ensures the exclusion of false recognitions. Each story was read aloud by the examiner, while the text remained in front of the participant who could read it again whenever necessary. After reading, the participants were asked the recognition question: Did anyone say something they should not have said? If participants answered "yes," they were asked the following four comprehension questions: (1) Who said something they should not have said? (2) Why should not he or she have said what he or she did? (3) Why did he or she say that? and (4) How did he or she feel? These questions are aimed to identify the character making the FP, judging the behavioural inadequacy, distinguishing intentional behaviour, and recognising emotions. In the case of a negative answer to the first question, or after the four FP questions described earlier, two open-ended control questions were asked to verify that the participant had understood the factual details of the story.

### Data handling

First, we describe how each task (the Prediction task, the Implicit Perspective-Taking task, the FP test, and finally, the Animal Game) was separately analysed to check that the expected results based on previous studies were individually replicated. Then, we describe the multiple regression analyses that we performed to explore whether the performance at the Animal Game could be moderated by either sensorimotor skills (indexed by the Prediction task) or inferential skills (indexed by the Implicit Perspective-Taking task and the FP test). The latter were named *Cognitive Bases of Joint Action (CBJA) analyses* as they aimed to provide indirect evidence of the role played by different cognitive skills (action prediction, perspective-taking, and theory of mind) in joint action.

**Prediction task data handling.** We measured the participant's Accuracy (ACC, i.e., proportion of correct responses) and RTs (measured in correct trials only). The ACC was at ceiling (70 errors in the whole sample, equal to 2.51% of the trials). With regard to the RTs, we calculated the individual mean in each experimental condition (Easy and Difficult) by excluding outlier values following the same rule reported earlier for data handling in the Animal Game (175 trials in the whole sample, equal to 6.29% of the trials). These data were log-transformed to approximate a normal distribution according to the Shapiro–Wilks test ( $p > .05$  in all conditions) and were analysed with a paired-sample *t*-test comparing the performance in Easy and Difficult trials as sanity-check. The log-transformed individual grand mean performance at the Prediction Task was taken as the *Action Prediction index (AP\_Index)* and entered the *CBJA analyses* reported below. The lower the

index, the better the participant's performance in the action Prediction task.

**Implicit perspective-taking task data handling.** As in the original papers (Samson et al., 2010; Surtees & Apperly, 2012), only trials in which the auditory description matched the picture ("yes" trials) were analysed (48 trials per participant). We measured the participant's Accuracy (ACC, i.e., proportion of correct responses) and RTs (measured in correct trials only). RTs were measured from the onset of the test picture showing the avatar in the room. The ACC was at ceiling (93 errors in the whole sample, equal to 3.34% of the trials, see Table 2). With regard to the RTs, we calculated the individual mean in each experimental condition (Self-Consistent, Self-Inconsistent, Other-Consistent, Other-Inconsistent) by excluding outlier values following the same rule reported above for data handling in the Animal Game (160 excluded trials in the whole sample, equal to 5.75% of the trials). These data were log-transformed to approximate a normal distribution according to the Shapiro–Wilks test ( $p > .05$  in all conditions) and then entered a preliminary within-subject ANOVA with Agent (Self/Other) and Consistency (Consistent/Inconsistent) as within-subject factors. As described in the original papers (Samson et al., 2010; Surtees & Apperly, 2012), a performance decay in the inconsistent as compared with the consistent condition is expected; in *Other-trials*, it would be evidence of egocentric bias, while in *Self-trials*, it would be evidence of implicit perspective-taking. In our study, the implicit perspective-taking was our variable of interest, and we thus calculated from the performance in Self-trials of each subject an *Implicit Perspective-taking index (PT\_Index)*, as follows:

$$PT\_Index = RTs \left[ \frac{(\text{Self-Inconsistent} - \text{Self-Consistent})}{(\text{Self-Inconsistent} + \text{Self-Consistent})} \right]$$

This index was log-transformed to approximate a normal distribution according to the Shapiro–Wilks test ( $p > .05$ ) after a linear transformation (having subtracted each value from 10) ensuring that all values had a positive sign (thus, allowing the log-transformation of the data). As stated earlier, this index is an estimate of the participants' tendency to involuntarily take into account the avatar's perspective, defined implicit perspective-taking in the original papers (Samson et al., 2010; Surtees & Apperly, 2012); the higher the index, the higher the implicit perspective-taking in the participant. The *PT\_Index* entered the *CBJA analyses* reported below.

**Faux Pas test data handling.** The FP test provides separate scores for the following aspects: (1) recognition of the presence or absence of a FP, (2) recognition of who did the FP, (3) recognition of the reasons why they said, (4) or did not say, something wrong, (5) a false belief score, and (6) a score regarding emotion recognition, and two scores

related to the control questions. Each score ranges from 0 to 10 (1 point attributed to each correct response). A total FP comprehension score can be also computed by summing up the first six scores.

To decide which specific score to enter in the *CBJA analysis* (see below), and whether all scores contributed to a unique cognitive dimension, i.e., *Theory of Mind (ToM)*, we performed a Principal Component Analysis (PCA). A total FP comprehension score was considered reliable (see the "Results" section), and it was then entered as the *Theory of Mind index (ToM\_Index)* in the *CBJA analyses* described below. The higher the index, the better the participant's scores indexing theory of mind abilities; this index was cubed to approximate a normal distribution.

**Animal game data handling.** We excluded from the analysis all trials in which the confederate made a false-start, that is, when she released the starting-button before hearing the auditory instructions (46 trials in the whole sample, equal to 0.62% of the trials), or in which she pressed the wrong button (51 trials in the whole sample, equal to 0.69% of the trials). The participant's false-starts (i.e., trials in which the participant released the starting-button before the go-signal) were also excluded from the analysis (180 trials in the whole sample, equal to 2.42% of the trials). In non-excluded trials, we measured the participant's Accuracy (ACC, i.e., proportion of correct responses) and RTs (i.e., the time-delay between the go-signal and the button-press as measured in correct trials only). The ACC could be considered at ceiling in both the JA and NI tasks (in the whole sample, only 48 errors occurred, equal to 0.67% of the trials). With regard to the RTs, we calculated the individual mean in each experimental condition by excluding outlier values using as threshold 1.5 times the interquartile distance (Tukey, 1977; 359 excluded trials in the whole sample, equal to 4.84% of the trials). These data were log-transformed to approximate a normal distribution according to the Shapiro–Wilks test ( $p > .05$  in all conditions).

As a preliminary analysis, a within-subject ANOVA was performed, with Task (JA/NI) and Action-Outcome Association (Co/Rev) as within-subject factors. We expected a Task by Association interaction effect with a selective performance decay in the JA-RevA condition (in line with previous studies; Sacheli, Arcangeli, & Paulesu, 2018; Sacheli, Meyer, et al., 2019); this would indicate that the participants find it more difficult to interact with a partner whose buttons function in a different way as compared with their own. This *JA-specific effect* was the one of interest in our study, and the one we planned to enter in *CBJA analyses* reported below; we aimed to test whether inter-individual differences in the effect-size of such an effect could be predicted by the inter-individual differences in either action prediction, visual perspective-taking, or theory of mind, as indexed by the *AP\_Index*, *PT\_Index*, and *ToM\_Index* described earlier. However,



**Table 2.** The table reports the group mean ( $\pm SD$ ) accuracy and raw response times values in each experimental condition of the three computer-based tasks.

Task	Accuracy	Response times (ms)
Prediction task		
Easy	0.98 $\pm$ 0.05	363.87 $\pm$ 117.04
Difficult	0.97 $\pm$ 0.04	416.51 $\pm$ 142.73
Implicit perspective-taking task		
Self-trials consistent	0.98 $\pm$ 0.05	859.15 $\pm$ 273.81
Self-trials inconsistent	0.94 $\pm$ 0.07	958.30 $\pm$ 287.72
Other-trials consistent	0.99 $\pm$ 0.03	767.88 $\pm$ 215.60
Other-trials inconsistent	0.96 $\pm$ 0.06	927.29 $\pm$ 279.76
Animal game		
JA-CoA	0.98 $\pm$ 0.03	780.04 $\pm$ 153.87
JA-RevA	0.99 $\pm$ 0.02	897.96 $\pm$ 198.83
NI-CoA	1.00 $\pm$ 0.01	778.33 $\pm$ 157.72
NI-RevA	1.00 $\pm$ 0.01	801.66 $\pm$ 165.71

CoA: coherent action-outcome association; JA: joint action; NI: non-interactive; RevA: reversed action-outcome association.

the ANOVA performed on the Animal Game data also showed a significant main effect of Action-Outcome Association (*Action-Outcome Association effect*, see the “Results” section), indicating that, independently of the task, the mean RTs in RevA trials were longer (i.e., performance was lower) than in the CoA trials (see the “Results” section). We then ran a second multiple regression to explore whether the three indexes reported earlier (the AP\_Index, PT\_Index, or ToM\_Index) could also moderate such an effect.

**Cognitive Bases of Joint Action (CBJA) analyses.** The CBJA analysis aimed to explore whether the effects of interests that emerged from the participants’ performance at the Animal Game could be moderated by the cognitive abilities described in Table 1 and indexed by the performance at (1) the Action Prediction task (AP\_index), (2) the Implicit Perspective-Taking task (PT\_index), and (3) the theory of mind FP test (ToM\_index).

We then ran two multiple regression analyses aimed to explore the possible presence of a linear association between the predictor indexes reported earlier (the AP\_Index, PT\_Index, or ToM\_Index, entered as independent variables in the multiple regressions) and two behavioural indexes of the significant effects emerged from the analysis of the Animal Game (the JA-specific effect and the Action-Outcome Association effect, entered as dependent variables in two separate multiple regression analyses). The Action-Outcome Association effect behavioural indexed was calculated as follows:

*Action - Outcome Association effect (Association effect)*

$$= RTs \left[ \frac{(RevA - CoA)}{(RevA + CoA)} \right]$$

The higher the index, the higher the performance decay associated with playing with a partner having a RevA as compared with the participant’s one. This index was also separately calculated per each task, and then the results used to calculate the JA-specific effect, as follows:

$$JA\text{-specific effect} = RTs [(JA\text{-Association effect}) - (NI\text{-Association effect})]$$

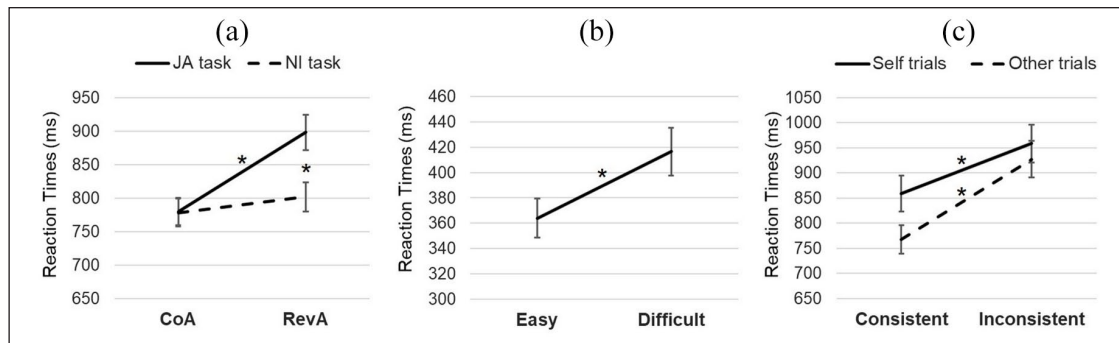
The higher the index, the higher the specific performance decay shown in the JA task (as compared with the NI task) associated with coordinating with a partner having a RevA as compared with the participant’s one.

Both these indexes were log-transformed to approximate a normal distribution according to the Shapiro–Wilks test ( $p > .05$ ) after a linear transformation (having subtracted each value from 10) ensuring that all values had a positive sign (thus allowing the log-transformation of the data). Before running the multiple regression analyses, we also checked that the three predictors (AP\_Index, PT\_Index, and ToM\_Index) did not correlate between each other, to exclude collinearity.

**Software.** The PCA was performed in the R environment (R Core Team, 2017); all other statistical analyses were performed in jamovi (version 1.6.23.0, <https://www.jamovi.org>, The jamovi project, 2020). All analyses were based upon an  $\alpha$ -level equal to .05. Bonferroni correction was performed when needed, e.g., to correct for multiple comparisons when performing post hoc tests.

## Results

For the sake of clarity, we report in Table 2 the group mean ( $\pm SD$ ) Accuracy and raw RTs values in the three experimental tasks. ACC was at ceiling in all tasks. Then, we



**Figure 3.** Results of the preliminary analyses in which the three computer-based experimental tasks were separately analysed. (a) The Task by Association interaction in the Animal Game, (b) the effect of Easy/Difficult trials in the Action Prediction task, and (c) the Self-/Other- by Consistency interaction effect in the Implicit Perspective-Taking task. \* indicates the significant comparisons.

**Table 3.** Factor loadings, communalities, and uniqueness of each separate index provided by FP test.

Indices of the FP test	Loadings	Communalities ( $h^2$ )	Uniqueness ( $u^2$ )
Recognition (presence/absence of a FP)	0.98	0.95	0.04
Identification ("Who did the FP?")	0.98	0.97	0.03
Reason why he/she said something wrong	0.97	0.94	0.06
Reason why he/she did not say something wrong	0.98	0.96	0.03
False belief	0.86	0.75	0.25
Emotion recognition	0.79	0.63	0.36

FP: Faux Pas.

report the analyses performed on each single task first, and finally the CBJA analyses.

### Action prediction task

The paired-sample  $t$ -test showed that the participants were faster in the Easy than the Difficult condition,  $t(57) = -8.40$ ,  $p < .001$ ,  $d = -1.10$ . This confirms that the participants were correctly performing the task and tried to guess the animal that would appear on the screen based on motor cues (so that the less the motor cues, as in the Difficult condition, the longer the RTs; see Figure 3).

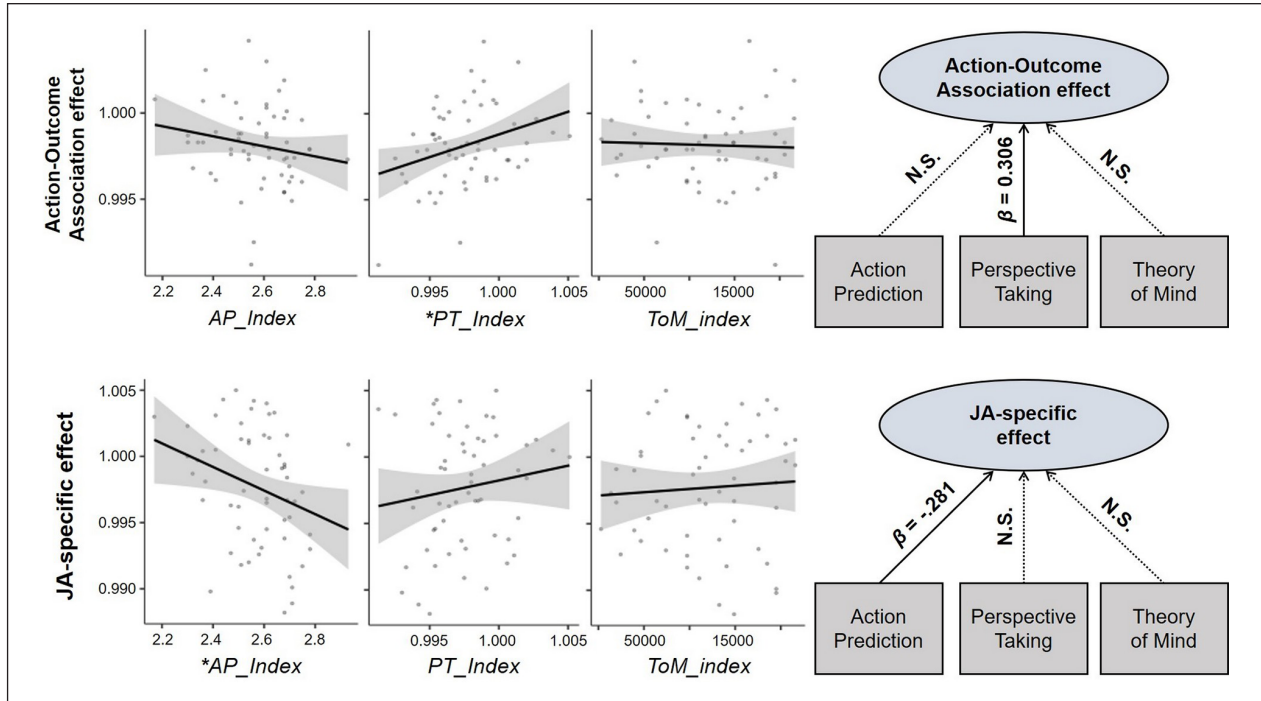
### Implicit perspective-taking task

The preliminary ANOVA showed a significant main effect of Agent,  $F(1, 57) = 25.73$ ,  $p < .001$ ,  $\eta_p^2 = .31$ , and Consistency,  $F(1, 57) = 123.28$ ,  $p < .001$ ,  $\eta_p^2 = .68$ , indicating that participants were faster in Other- than Self-trials (Self  $908.72 \pm 273.68$  ms; Other  $847.58 \pm 232.93$  ms) and in Consistent than Inconsistent trials (Consistent  $813.52 \pm 235.98$  ms; Inconsistent  $942.79 \pm 274.12$  ms). As in the original paper (Samson et al., 2010), the results also showed a significant Agent  $\times$  Consistency interaction,  $F(1, 57) = 8.56$ ,  $p = .005$ ,  $\eta_p^2 = .13$ , indicating that, while the difference between Consistent and Inconsistent trials was equally present in Self- and Other-trials (both

$p_{corr} < .001$ ), the RTs in Self-trials were slower than in Other-trials in the Consistent condition only ( $p_{corr} < .001$ ). As stated in the "Methods" section, the presence of a significant Consistency effect in Self-trials is taken as evidence of Implicit Perspective-Taking in our participants (see Figure 3).

### Faux Pas test

The PCA run on the six indices extracted by the FP test was performed by using the *principal* routine of the "stats" R package (Revelle & Revelle, 2015) and applying an orthogonal rotation (varimax). The Kaiser–Meyer–Olkin value confirmed the sampling adequacy ( $KMO = 0.92$ ) and all  $KMO$  for individual indices were  $> 0.8$  (highly above 0.5, i.e., the threshold indicated by Kaiser, 1974). Bartlett's test of sphericity,  $\chi^2(15) = 577.81$ ,  $p < .001$ , indicated that the correlations between variables were different from 0. On the basis of the eigenvalue  $> 1$  and of the screen-plot exploration, a unique component was extracted by the PCA (eigenvalue = 5.2; see factor loadings for each variable in Table 3), which explained the 86% of variance. In the light of these results, the separate scores provided by the FP test could be pooled together in a unidimensional measure of ToM. Accordingly, the total FP score (mean score  $47.22 \pm 10.20$ ) was used as *ToM\_Index* in the CBJA analyses.



**Figure 4.** The plots illustrate the results of the Cognitive Bases of Joint Action (CBJA) analyses. While the PT\_index moderated the strength of the Action-Outcome Association effect, the AP\_index was a specific moderator of the JA-specific effect. Significant effects are marked with an (\*). The ToM\_index was not significant as moderator in either analysis, as also confirmed by the Bayesian correlation analysis.

### Animal game (Joint action and non-interactive tasks)

The results showed a significant main effect of Task,  $F(1, 57)=15.0$ ,  $p<.001$ ,  $\eta_p^2=.21$ , and Action-Outcome Association,  $F(1, 57)=33.5$ ,  $p<.001$ ,  $\eta_p^2=.37$ , indicating that participants were faster in the NI than the JA task (NI  $790.00 \pm 152.72$  ms; JA  $839.00 \pm 161.93$  ms) and in trials where the partner's Action-Outcome Association was Coherent than Reversed (Coherent  $779.19 \pm 147.66$  ms, Reversed  $849.81 \pm 167.36$  ms). As expected, the results also showed a Task  $\times$  Action-Outcome Association significant interaction ( $F(1, 57)=15.3$ ,  $p<.001$ ,  $\eta_p^2=.21$ ) indicating that the Action-Outcome Association effect was significant in the JA ( $p_{corr}<.001$ ) but not the NI task ( $p_{corr}=.88$ ). The RTs in the JA-RevA condition were slower than in all other experimental conditions (all  $p_{s_{corr}}<.001$ , see Figure 3). These results indicate that, as expected based on the previous studies, the interaction with a partner whose buttons functioned in a way different from the participant's one was more difficult only in the JA task, when the participant had to respond to the partner's action, and not in the NI task, where the partner's action was irrelevant to the participant's task.

### Cognitive Bases of Joint Action (CBJA) analyses

The three predictors did not correlate between each other, thus excluding collinearity (all  $r_s < .16$  and all  $p_s > .2$ ). As

explained in the “Methods” section, we calculated two indexes of the effects of interest that emerged from the analysis of the Animal Game task, namely, (1) the *JA-specific effect*, indexing the individuals' specific performance decay shown in the JA task (as compared with the NI task) and associated with coordinating with a partner having a RevA as compared with the participant's one and (2) the *Action-Outcome Association effect*, indexing the performance decay associated with playing with a partner having a RevA as compared with the participant's one, independently of the task (JA or NI).

The multiple regression analysis having the *JA-specific effect* as dependent variable revealed a significant effect of the AP\_Index ( $t(56)=-2.15$ ,  $p=.036$ ,  $\beta=-.281$ ), while the other two predictors were not significant ( $p_s > .34$ ). The direction of the effect indicates that the lower the AP\_index (i.e., the better the performance at the Action Prediction task), the higher the *JA-specific effect* (i.e., the stronger it was the participant's slowdowns in RTs in the Reversed as compared with CoA trials in the JA as compared with the NI task; see Figure 4, lower panel).

The multiple regression analysis having the *Action-Outcome Association effect* as dependent variable revealed a significant effect of the PT\_Index,  $t(56)=2.40$ ,  $p=.020$ ,  $\beta=.306$ , while the other two predictors were not significant ( $p_s > .19$ ). The direction of the effect indicates that the higher the PT\_index (i.e., the stronger the implicit perspective-taking in the participant), the higher the

participant's overall slowdown in RTs in the Reversed as compared with the CoA trials, independently of the task (see Figure 4, upper panel).

To ensure that the absence of any significant result in the analyses having the *ToM\_index* as predictor was not due to a lack of power, we performed a *Bayesian correlation analysis*. Bayesian Factor (BF) is a statistical metric that quantifies the strength of evidence that the data provide in favour of the alternative hypothesis relative to the null hypothesis; a BF10 higher than 3 indicates substantial evidence in favour of the alternative hypothesis, whereas a BF10 lower than 0.3 indicates substantial evidence in favour of the null hypothesis (Jarosz & Wiley, 2014). We thus ran a Bayesian correlation analysis having as an alternative hypothesis the presence of a correlation between the *ToM\_index* and (1) the *Action-Outcome Association effect* or (2) the *JA-specific effect* in the Animal Game, and the absence of correlation as a null hypothesis. The results indicated evidence in favour of the null hypothesis both for the first (BF=0.171) and the second (BF=0.184) correlation.

## Discussion

This study investigates the socio-cognitive bases of small-scale interactions, which are the building blocks for joint action, by testing the role of inferential and predictive motor mechanisms in moderating adult individuals' performance at joint action tasks. Clarifying what socio-cognitive processes influence the ability to perform joint actions might contribute to the theoretical debate in the literature, which opposes two contrasting views on what motor interactions require (see Table 1).

On one hand, it has been suggested that small-scale interactions are grounded on shared intentions (Bratman, 1992, 1999, 2014; Carpenter, 2009), that is, the knowledge, on the part of each agent involved in the interaction, that they all know that they each intend to do something together. The possibility to represent shared intentions requires, by definition, theory of mind, that is, the ability to represent others' mental states. On the other hand, joint action abilities might be based on the capacity to represent shared goals (Butterfill, 2012) and to apply the ensuing sensorimotor predictive processes to ensure mutual adaptation (Candidi et al., 2015; Pesquita et al., 2018; Sacheli, Arcangeli, & Paulesu, 2018).

Thus, we conceived an experimental design that allowed us to measure in a sample of 58 healthy, young adult individuals (age  $22.31 \pm 3.12$ ), the performance on a joint action task and also three further tasks, namely (1) an action prediction task, (2) a perspective-taking task, and (3) a theory of mind test. To measure joint action performance, we built up a novel Animal Game that reflected the rationale of previous published joint action tasks applied in adults and children (Sacheli, Arcangeli, & Paulesu, 2018, 2019). The task required the participants to coordinate with a partner (who

was the experimenter's confederate) while pressing one of two buttons associated with the appearance of cartoon-like animals on the screen. This task was performed both in an interactive (JA) and in a perceptually matched, NI situation. The crucial experimental manipulation regarded the partner's button-animal association inversion in 50% of the trials (RevA condition). Importantly, and in contrast with previous studies, our participants were fully aware of this manipulation. The RevA condition implied that, in the JA task, the participants had to bear in mind the specific partner's button functioning if they wanted to anticipate what animal would appear on the screen by observing the partner's movements, to prepare their own response in advance. We expected this attempt to predict the partner's animal to take place only in the JA and not in the NI task, as in the NI task the partner's animal is irrelevant.

In our task, these predictive attempts would lead to performance decay, because the partner's reversed button-animal association would conflict with the association stored in the partner's motor system. In other words, stronger predictions (necessarily based on one's own action-effect association) would improve the performance in the CoA trials (as the participants can anticipate which animal will appear on the screen and prepare a response in advance) and impair it in the RevA trials (due to "wrong" predictions based on one's action-effect association), thus generating a stronger effect of Action-Outcome Association (Reversed > Coherent). This effect should be selective for the JA task, as in the NI task, there is no need for prediction and anticipation as the partner's animal is irrelevant for the participants' task. On the contrary, better perspective-taking could reduce such an effect, enabling participants to deal with the conflicting representations regarding one's and the partner's action-effect associations.

Our results showed a performance decay in the Reversed as compared with the CoA condition, which was specific for the JA task, as confirmed by the Task by Action-Outcome Association significant interaction. Importantly, they also indicated that this performance decay is moderated by the individual's action prediction abilities. The analyses also showed a significant main effect of the Action-Outcome Association in the Animal Game, and a significant role of perspective-taking in moderating this effect, a result that is further discussed below.

### *The role of action prediction in moderating the task by action-outcome association interaction*

With regard to the interaction effect, we calculated an index measuring the effect-size of such interaction effect in each participant and tested whether the strength of this effect could be moderated by indexes of the participants' abilities in (1) action prediction, (2) perspective-taking, and (3) theory of mind, as measured by the other tasks included in the experiment. The rationale of this design



was as follows. As the performance decay in the Reversed as compared with the CoA condition can be interpreted as the results of *conflicting representations* regarding the agent's and the partner's button functioning, it might well be that the ability to represent others' perspectives and mental states might moderate such an effect. Perceptive-taking and theory of mind skills are indeed interpreted as the ability to keep track of conflicting information regarding the self and the others (Frick & Baumeler, 2017). However, if the cause of the decay is the participants' active use of these representations in attempting to *predict which animal* will appear on the screen by observing the partner's moves, then, action prediction abilities might play a major role.

The results were in favour of the second hypothesis. Indeed, action prediction abilities were good predictors of the selective performance decay in the RevA condition that the participants showed in the JA task. These results are in line with the stand-points suggesting that motor predictive processes are crucial in JA (Knoblich & Jordan, 2003; Sebanz & Knoblich, 2009; Vesper et al., 2013) and allow taking into account and anticipating the sensory consequence of a partner's action (Pfister et al., 2014; Sacheli, Arcangeli, & Paulesu, 2018), possibly thought the activity of fronto-parietal areas (Era et al., 2018; Hadley et al., 2015; Sacheli, Candidi, et al., 2015; Sacheli, Tieri, et al., 2018; Sacheli, Verga, et al., 2019). Future studies may directly test the involvement of these fronto-parietal sensorimotor networks in our tasks with neurophysiological techniques.

It is worth noting here that our interpretation of results is apparently contradicting; how can showing that better action prediction causes a greater *decay* in joint action performance really support the hypothesis that these action prediction abilities *ground* joint action abilities? This apparent paradox can be resolved by noting that the performance decay is a consequence of a highly artificial intervention—reversing the association between a partner's actions and their outcomes—which has been designed to expose the degree to which participants distinguish joint actions from parallel but merely individual actions (Sacheli, Arcangeli, & Paulesu, 2018). This is why we can interpret greater decay in performance as evidence for stronger engagement in joint action.

Our results also provide evidence that the recruitment of predictive processes depends on the representation of a shared goal (Butterfill, 2012), for such processes are not recruited in the absence of a shared goal (in the NI task). The specificity of the association between predictive abilities and the JA (but not the NI) task also ensures that our results were not influenced by the similarity between the visual stimuli and set-up used in the Prediction task and the Animal Game, as this similarity was present in both the NI and the JA tasks, but the association with predictive abilities was specific for the latter.

### *The role of perspective-taking in moderating the main effect of association*

Importantly, we are not suggesting that inferential processes play no role in interaction. We do think that inferential processes might affect performance in social exchanges (see Curioni & Sacheli, 2019), and our data seem to support this hypothesis. As a matter of fact, although our task was a very easy motor task, perspective-taking did moderate the participants' performance and specifically the strength of the main effect of Action-Outcome Association, confirming the general role of perspective-taking in the social context. Obviously, we cannot exclude that by parametrising the task demands in both the perspective-taking and the joint action tasks the results may have shown a more distinctive association between more sophisticated perspective-taking and JA situations. However, as far as the present results are concerned, they clearly indicate an association between perspective-taking and both the JA and the NI task. Thus, our data suggest that perspective-taking and the ensuing possibility to keep in mind and possibly integrate conflicting information regarding the self and the other individuals is indeed foundational of all social exchanges, independently of whether we need to coordinate with others or not. By contrast, action prediction abilities seem to be selectively recruited in JA and to moderate the participants' performance only in this context.

These results may contribute to the current debates not only on motor interaction but also on the methodological approaches used to study social cognition. Indeed, we show that the interactivity of the social context can affect how much different cognitive abilities (e.g., action prediction and perspective-taking) are involved. Our data suggest that perspective-taking plays a role in any social situations in which we cannot avoid taking into account the others' behaviour, even if it is simply because we have to take turns (such as in the NI task resented here); these situations include also (but not only) joint actions guided by a shared goal.

### *Conclusions and future directions*

Our results may seem surprising in the light of previous studies suggesting that social skills might play a significant role in modulating the participants' motor behaviours in what could be interpreted as motor interactions (Brass et al., 2009; Sowden & Catmur, 2015; Spengler et al., 2009). Furthermore, a developmental study has shown that theory of mind abilities moderate 4-year-old children's ability to represent another agent's (irrelevant) task (Milward et al., 2017). However, these previous studies applied motor tasks that did not involve any common goal shared between the interactive agents. We suggest that it is the presence of a shared goal what makes sensorimotor predictive mechanisms critical for moderating the interaction performance (see also Sacheli, Aglioti, & Candidi, 2015).

Of course, it might well be that measuring more empathy-related than purely cognitive social skills could lead to different results (see, for instance, Novembre et al., 2019). Moreover, interactions of a different kind and requiring different task demands, e.g., those involving more complex and temporally extended tasks, or those mediated by verbal communication, might require different cognitive processes and recruit inferential abilities like theory of mind to a greater extent. For instance, the strategic use of coordination smoothers is reduced in individuals with high autistic traits (Curioni et al., 2017), thus suggesting that it might be related to the agent's sophisticated social skills (Curioni & Sacheli, 2019). Nevertheless, saying that more efficient interactive partners possibly capitalise on their social skills (like theory of mind) to interact more flexibly is not the same as saying that motor interactions *necessarily require* inferential socio-cognitive skills. We have not shown that small-scale social interactions only ever require sensorimotor predictive mechanisms; but our results suggest that the latter might be especially relevant when interaction is guided by a shared goal, while social skills might have a broader influence on any form of social exchange (either requiring a shared goal or not).

This empirical evidence makes coherent a speculation about development, which suggests a way the theoretical debate illustrated in Table 1 might be partly reconciled. It might well be that the repeated experience of interactions with others, scaffolded by the representation of shared goals and by the motor predictions they trigger (see also Krogh-Jespersen et al., 2020), enables children to fine-tune their social skills to make them of use as the interaction unfolds, as suggested by constructivist views of social development (Carpendale & Lewis, 2004; Luyten & Fonagy, 2015; Sodian et al., 2020; Tomasello, 2018; see evidence from Jin et al., 2018). As a consequence, small-scale interactions may not necessarily depend on high-level social cognition, but rather represent the scaffolding condition for their development. This speculation could be addressed by future studies.

### Acknowledgements

The authors would like to dedicate this paper to the fond memory of Dr Enea Francesco Pavone, founder of BrainTrends, for his essential assistance in the development of the experimental set-up.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Lucia Maria Sacheli  <https://orcid.org/0000-0003-2608-708X>

### Data accessibility statement



The data from the present experiment are publicly available at the Open Science Framework website: [osf.io/etnfx](https://osf.io/etnfx).

### References

- Adolph, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, 11, 231–239.
- Avenanti, A., Annella, L., Candidi, M., Urgesi, C., & Aglioti, S. M. (2013). Compensatory plasticity in the action observation network: Virtual lesions of STS enhance anticipatory simulation of seen actions. *Cerebral Cortex*, 23(3), 570–580.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42(2), 241–251. <http://www.ncbi.nlm.nih.gov/pubmed/11280420>
- Bekkering, H., De Bruijn, E. R. A., Cuijpers, R. H., Newman-Norlund, R., Van Schie, H. T., & Meulenbroek, R. (2009). Joint action: Neurocognitive mechanisms supporting human interaction. *Topics in Cognitive Science*, 1(2), 340–352. <https://doi.org/10.1111/j.1756-8765.2009.01023.x>
- Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528), 2359–2367. <https://doi.org/10.1098/rstb.2009.0066>
- Bratman, M. (1992). Shared cooperative activity. *The Philosophical Review*, 101(2), 327–341.
- Bratman, M. (1999). *Faces of intention: Selected essays on intention and agency*. Cambridge University Press.
- Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford University Press.
- Brownell, C. A. (2011). Early developments in joint action. *Review of Philosophy and Psychology*, 2, 193–211. <https://doi.org/10.1007/s13164-011-0056-1>
- Brownell, C. A., Ramani, G. B., & Zerwas, S. (2006). Becoming a social partner with peers: Cooperation and social understanding in one- and two-year-olds. *Child Development*, 77(4), 803–821. <https://doi.org/10.1111/j.1467-8624.2006.00904.x>
- Butterfill, S. (2012). Joint action and development. *Philosophical Quarterly*, 62(246), 23–47. <https://doi.org/10.1111/j.1467-9213.2011.00005.x>
- Candidi, M., Sacheli, L. M., & Aglioti, S. M. (2015). From muscles synergies and individual goals to interpersonal synergies and shared goals: Mirror neurons and interpersonal action hierarchies. Comment on “Grasping synergies: A motor-control approach to the mirror neuron mechanism” by D’Ausilio et al. *Physics of Life Reviews*, 12, 126–128. <https://doi.org/10.1016/j.plrev.2015.01.023>
- Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children’s social understanding within social interaction. *Behavioural and*

- Brain Sciences*, 27(1), 79–151. <https://doi.org/10.1017/s0140525x04000032>
- Carpenter, M. (2009). Just how joint is joint action in infancy? *Topics in Cognitive Science*, 1(2), 380–392. <https://doi.org/10.1111/j.1756-8765.2009.01026.x>
- Curioni, A., Knoblich, G. K., Sebanz, N., & Sacheli, L. M. (2020). The engaging nature of interactive gestures. *Plos one*, 15(4), e0232128.
- 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 behaviour. *Molecular Autism*, 8(1), Article 23. <https://doi.org/10.1186/s13229-017-0141-0>
- Curioni, A., & Sacheli, L. M. (2019). The role of social learning and socio-cognitive skills in sensorimotor communication: Comment on “The body talks: Sensorimotor communication and its brain and kinematic signatures” by Pezzulo et al. *Physics of Life Reviews*, 28, 24–27. <https://doi.org/10.1016/j.plrev.2019.01.021>
- Era, V., Candidi, M., Gandolfo, M., Sacheli, L. M., & Aglioti, S. M. (2018). Inhibition of left anterior intraparietal sulcus shows that mutual adjustment marks dyadic joint-actions in humans. *Social Cognitive and Affective Neuroscience*, 13(5), 492–500. <https://doi.org/10.1093/scan/nsy022>
- Finisguerra, A., Amoroso, L., & Urgesi, C. (2020). Beyond automatic motor mapping: New insights into top-down modulations on action perception. In N. Noceti, A. Sciutti & F. Rea (Eds.), *Modelling human motion* (pp. 33–51). Springer.
- Frick, A., & Baumeler, D. (2017). The relation between spatial perspective taking and inhibitory control in 6-year-old children. *Psychological Research*, 81(4), 730–739. <https://doi.org/10.1007/s00426-016-0785-y>
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1431), 459–473. <https://doi.org/10.1098/rstb.2002.1218>
- Furlanetto, T., Becchio, C., Samson, D., & Apperly, I. (2016). Altercentric interference in level 1 visual perspective taking reflects the ascription of mental states, not submentalizing. *Journal of Experimental Psychology: Human Perception and Performance*, 42(2), 158–163.
- Hadley, L. V., Novembre, G., Keller, P. E., & Pickering, M. J. (2015). Causal role of motor simulation in turn-taking behaviour. *Journal of Neuroscience*, 35(50), 16516–16520. <https://doi.org/10.1523/JNEUROSCI.1850-15.2015>
- Heyes, C. M. (2014). Submentalizing: I’m not really reading your mind. *Psychological Science*, 9, 131–143.
- Holland, C., Shin, S. M., & Phillips, J. (2021). *Do you see what I see? A meta-analysis of the Dot Perspective Task*. Proceedings of the Annual Meeting of the Cognitive Science Society, no. 43. <https://escholarship.org/uc/item/7cs5r2xq>
- The jamovi project. (2020). *jamovi (Version 1.2)* [Computer Software]. <https://www.jamovi.org>
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, 7(1), Article 2.
- Jin, X., Li, P., He, J., & Shen, M. (2018). How you act matters: The impact of coordination on 4-year-old children’s reasoning about diverse desires. *Journal of Experimental Child Psychology*, 176, 13–25. <https://doi.org/10.1016/j.jecp.2018.07.002>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36.
- Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, 15(8), 352–357.
- Knoblich, G., & Jordan, J. S. (2003). Action coordination in groups and individuals: Learning anticipatory control. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 29(5), 1006–1016. <https://doi.org/10.1037/0278-7393.29.5.1006>
- Krogh-Jespersen, S., Henderson, A. M. E., & Woodward, A. L. (2020). Let’s get it together: Infants generate visual predictions based on collaborative goals. *Infant Behaviour and Development*, 59, Article 101446. <https://doi.org/10.1016/j.infbeh.2020.101446>
- Luyten, P., & Fonagy, P. (2015). The neurobiology of mentalizing. *Personality Disorders*, 6(4), 366–379. <https://doi.org/10.1109/TCSET.2006.4404627>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behaviour Research Methods*, 44(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Meyer, M., Bekkering, H., Haartsen, R., Stapel, J. C., & Hunnius, S. (2015). The role of action prediction and inhibitory control for joint action coordination in toddlers. *Journal of Experimental Child Psychology*, 139, 203–220. <https://doi.org/10.1016/j.jecp.2015.06.005>
- Meyer, M., Bekkering, H., Paulus, M., & Hunnius, S. (2010). Joint action coordination in 2½- and 3-year-old children. *Frontiers in Human Neuroscience*, 4, Article 220. <https://doi.org/10.3389/fnhum.2010.00220>
- Milward, S. J., Kita, S., & Apperly, I. A. (2017). Individual differences in children’s corepresentation of self and other in joint action. *Child Development*, 88(3), 964–978. <https://doi.org/10.1111/cdev.12693>
- Newcombe, N. (1989). The development of spatial perspective taking. In H. W. Reese (Ed.), *Advances in child development and behaviour* (pp. 203–247). Academic Press.
- Novembre, G., Mitsopoulos, Z., & Keller, P. E. (2019). Empathic perspective taking promotes interpersonal coordination through music. *Scientific Reports*, 9(1), Article 12255. <https://doi.org/10.1038/s41598-019-48556-9>
- Pavlidou, A., Gallagher, M., Lopez, C., & Ferrè, E. R. (2019). Let’s share our perspectives, but only if our body postures match. *Cortex*, 119, 575–579.
- Pesquita, A., Whitwell, R. L., & Enns, J. T. (2018). Predictive joint-action model: A hierarchical predictive approach to human cooperation. *Psychonomic Bulletin and Review*, 25(5), 1751–1769. <https://doi.org/10.3758/s13423-017-1390-6>
- Pezzulo, G., Iodice, P., Donnarumma, F., Dindo, H., & Knoblich, G. (2017). Avoiding accidents at the champagne reception: A study of joint lifting and balancing. *Psychological Science*, 28(3), 338–345. <https://doi.org/10.1177/0956797616683015>
- Pfister, R., Dolk, T., Prinz, W., & Kunde, W. (2014). Joint response-effect compatibility. *Psychonomic Bulletin and Review*, 21(3), 817–822. <https://doi.org/10.3758/s13423-013-0528-7>



- R Core Team. (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. <https://www.R-project.org/>.
- Revelle, W., & Revelle, M. W. (2015). *Package "psych."* *The comprehensive R archive network*, 337, 338.
- Sacheli, L. M., Aglioti, S. M., & Candidi, M. (2015). Social cues to joint actions: The role of shared goals. *Frontiers in Psychology*, 6, Article 1034. <https://doi.org/10.3389/fpsyg.2015.01034>
- Sacheli, L. M., Arcangeli, E., & Paulesu, E. (2018). Evidence for a dyadic motor plan in joint action. *Scientific Reports*, 8, Article 5027. <https://doi.org/10.1038/s41598-018-23275-9>
- Sacheli, L. M., Candidi, M., Era, V., & Aglioti, S. M. (2015). Causative role of left aIPS in coding shared goals during human-avatar complementary joint actions. *Nature Communications*, 6, Article 7544. <https://doi.org/10.1038/ncomms8544>
- Sacheli, L. M., Meyer, M., Hartstra, E., Bekkering, H., & Hunnius, S. (2019). How preschoolers and adults represent their joint action partner's behaviour. *Psychological Research*, 83(5), 863–877. <https://doi.org/10.1007/s00426-017-0929-8>
- Sacheli, L. M., Musco, M., Zazzera, E., & Paulesu, E. (2021). Mechanisms for mutual support in motor interactions. *Scientific Reports*, 11(1), Article 3060. <https://doi.org/10.1038/s41598-021-82138-y>
- Sacheli, L. M., Tieri, G., Aglioti, S. M., & Candidi, M. (2018). Transitory inhibition of the left anterior intraparietal sulcus impairs joint actions: A continuous theta-burst stimulation study. *Journal of Cognitive Neuroscience*, 30(5), 737–751. [https://doi.org/10.1162/jocn\\_a\\_01227](https://doi.org/10.1162/jocn_a_01227)
- Sacheli, L. M., Verga, C., Arcangeli, E., Banfi, G., Tettamanti, M., & Paulesu, E. (2019). How task interactivity shapes action observation. *Cerebral Cortex*, 29(12), 5302–5314.
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1255–1266. <https://doi.org/10.1037/a0018729>
- Santisteban, I., Catmur, C., Hopkins, S., Bird, G., & Heyes, C. M. (2014). Avatars and arrows: Implicit mentalizing or domain general processing? *Journal of Experimental Psychology: Human Perception and Performance*, 40, 929–937.
- Sartori, L., Becchio, C., Bulgheroni, M., & Castiello, U. (2009). Modulation of the action control system by social intention: unexpected social requests override preplanned action. *Journal of Experimental Psychology: Human Perception and Performance*, 35(5), 1490.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, 10(2), 70–76.
- Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, 1(2), 353–367. <https://doi.org/10.1111/j.1756-8765.2009.01024.x>
- Sinigaglia, C., & Butterfill, S. A. (2020). Motor representation and knowledge of skilled action. In E. Fridland & C. Pavese (Eds.), *The Routledge handbook of philosophy of skill and expertise* (pp. 292–305). Routledge.
- Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How does children's theory of mind become explicit? A review of longitudinal findings. *Child Development Perspectives*, 14(3), 171–177. <https://doi.org/10.1111/cdep.12381>
- Sowden, S., & Catmur, C. (2015). The role of the right temporoparietal junction in the control of imitation. *Cerebral Cortex*, 25(4), 1107–1113. <https://doi.org/10.1093/cercor/bht306>
- Spengler, S., Von Cramon, D. Y., & Brass, M. (2009). Control of shared representations relies on key processes involved in mental state attribution. *Human Brain Mapping*, 30(11), 3704–3718. <https://doi.org/10.1002/hbm.20800>
- Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10(5), 640–656.
- Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, 83(2), 452–460. <https://doi.org/10.1111/j.1467-8624.2011.01730.x>
- Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences of the United States of America*, 115(34), 8491–8498. <https://doi.org/10.1073/pnas.1804761115>
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioural and Brain Sciences*, 28(5), 675–691; discussion 691–735. <https://doi.org/10.1017/S0140525X05000129>
- Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Addison-Wesley.
- Urgesi, C., Maieron, M., Avenanti, A., Tidoni, E., Fabbro, F., & Aglioti, S. M. (2010). Simulating the future of actions in the human corticospinal system. *Cerebral Cortex*, 20(11), 2511–2521.
- Urgesi, C., Moro, V., Candidi, M., & Aglioti, S. M. (2006). Mapping implied body actions in the human motor system. *Journal of Neuroscience*, 26(30), 7942–7949.
- Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2013). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 48–61. <https://doi.org/10.1037/a0028066>
- Ward, E., Ganis, G., & Bach, P. (2019). Spontaneous vicarious perception of the content of another's visual perspective. *Current Biology*, 29(5), 874–880.e4. <https://doi.org/10.1016/j.cub.2019.01.046>
- Warneken, F., Gräfenhain, M., & Tomasello, M. (2012). Collaborative partner or social tool? New evidence for young children's understanding of joint intentions in collaborative activities. *Developmental Science*, 15(1), 54–61. <https://doi.org/10.1111/j.1467-7687.2011.01107.x>
- Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, 11(3), 271–294. <https://doi.org/10.1111/j.1532-7078.2007.tb00227.x>