



**HAL**  
open science

## Addressing joint action challenges in HRI: Insights from psychology and philosophy

Kathleen Belhassein, Víctor Fernández-Castro, Amandine Mayima, Aurélie Clodic, Elisabeth Pacherie, Michèle Guidetti, Rachid Alami, Hélène Cochet

### ► To cite this version:

Kathleen Belhassein, Víctor Fernández-Castro, Amandine Mayima, Aurélie Clodic, Elisabeth Pacherie, et al.. Addressing joint action challenges in HRI: Insights from psychology and philosophy. *Acta Psychologica*, 2022, 222, 10.1016/j.actpsy.2021.103476 . hal-03505047

**HAL Id: hal-03505047**

**<https://hal.science/hal-03505047>**

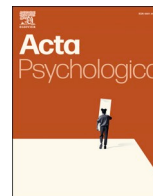
Submitted on 10 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



## Addressing joint action challenges in HRI: Insights from psychology and philosophy

Kathleen Belhassein<sup>a,b</sup>, Víctor Fernández-Castro<sup>c,\*</sup>, Amandine Mayima<sup>b</sup>, Aurélie Clodic<sup>b</sup>, Elisabeth Pacherie<sup>d</sup>, Michèle Guidetti<sup>a</sup>, Rachid Alami<sup>b</sup>, Hélène Cochet<sup>a</sup>

<sup>a</sup> CLLE, UMR5263, Toulouse University, CNRS, UT2J, France

<sup>b</sup> LAAS-CNRS, UPR8001, Toulouse University, CNRS, France

<sup>c</sup> Filo-Lab Unit, University of Granada, Spain

<sup>d</sup> Institut Jean Nicod, UMR8129, CNRS, DEC, ENS, PSL University, France

### ARTICLE INFO

#### Keywords:

HRI  
Joint action  
Communication  
Mutual recognition  
Philosophy  
Psychology

### ABSTRACT

The vast expansion of research in human-robot interactions (HRI) these last decades has been accompanied by the design of increasingly skilled robots for engaging in joint actions with humans. However, these advances have encountered significant challenges to ensure fluent interactions and sustain human motivation through the different steps of joint action. After exploring current literature on joint action in HRI, leading to a more precise definition of these challenges, the present article proposes some perspectives borrowed from psychology and philosophy showing the key role of communication in human interactions. From mutual recognition between individuals to the expression of commitment and social expectations, we argue that communicative cues can facilitate coordination, prediction, and motivation in the context of joint action. The description of several notions thus suggests that some communicative capacities can be implemented in the context of joint action for HRI, leading to an integrated perspective of robotic communication.

### 1. Introduction

In the following decades, human societies will witness a pervasive use of robotic agents in all contexts of public and private social interactions. Social robotics is currently producing and designing robotic agents with use in numerous contexts like game companion (Sanghvi et al., 2011), education and therapy (Brage et al., 2018; McGlynn et al., 2014), or services (Kanda et al., 2010). For continuing these advances, social robotics needs to successfully design robots able to engage with humans, so they can collaborate on shared activities which require high levels of coordination. This need explains the fast expansion of the field of human-robot interaction (HRI), which attempts to develop different avenues for enabling robots to encounter social interactions. As part of this expansion, HRI research has taken inspiration from some important findings in psychology, philosophy of mind, and neuroscience to provide robotic agents with the necessary cognitive capabilities for achieving joint actions (Giger et al., 2019; Thomaz et al., 2016).

This approach in social robotics focuses on equipping robots with devices based on human psychological mechanisms underlying shared

activities. Some of the mechanisms that roboticists have attempted to design include theory of mind, emotional recognition, or human-aware navigation (see Thomaz et al. (2016) for a review). Despite these advances, research in HRI suggests that equipping robots with social skills can sometimes, rather counterintuitively, undermine user experience and hinder the interaction between humans and robots (Giger et al., 2019; Sciutti et al., 2018). For instance, robot's human-like appearance or personality can be perceived as deceptive (Vandemeulebroucke et al., 2018): human robot interactions may be undermined by the novelty effect or the expectation gap between what people believe about the robot—especially when people did not have contact with robots and their expectations are infected by popular culture (see e.g., Sandoval et al., 2014)—and the actual competence of the robots (see Kwon et al., 2016, also Section 2.3.2). Furthermore, some of the robot's attributes or behaviors, like for example head-nodding (Thepsoonthorn et al., 2021), may trigger attributions of minds (Gray & Wegner, 2012) leading to a feeling of strangeness or unfamiliarity (known as the Uncanny valley effect), which can impact the humans' levels of trust toward the robot (Lewis et al., 2018). These negative effects may be increased with the

\* Corresponding author at: Departamento de Filosofía I Edificio de la Facultad de Psicología, Campus de la Cartuja, 18011 Granada, Spain.

E-mail address: [vfernandezcastro@ugr.es](mailto:vfernandezcastro@ugr.es) (V. Fernández-Castro).

implementation of certain social capacities and behaviors, especially when they are implemented in isolation. For instance, [Thepsoonthorn et al. \(2021\)](#) have found that the feeling of uncanniness related to head nodding does not appear when the behavior is accompanied with a manual gesture. Also, [Riek et al. \(2010\)](#) found that certain types of cooperative behaviors are more effective than others, but also that negative attitudes toward robots are strongly correlated with a decreased ability in decoding human gestures. The choice to focus on some socio-cognitive mechanisms in isolation usually goes with expertise in the field, given that specialization usually requires researchers to “dissect” some processes or situations. This appears necessary to gain some expertise, however, it also raises important questions for social robotics and its attempt to equip robots with the capacity to carry out joint action with humans.<sup>1</sup> Are we gaining expertise at the expense of considering the “broad picture”? Are we missing something when focusing on specific socio-cognitive capacities? Are there different strategies to explore to ensure fluency to interact with the robot? Besides, are there general strategies to influence human attitudes during HRI?

The aim of this paper is twofold. First, we argue that the sub-optimal or even negative effects of considering and designing specific capacities in isolation can be overcome if we focus on communicative capacities seen in human interactions, which can be implemented in the context of joint action for HRI. Second, we review psychological and philosophical literature that explores these human communicative capabilities to provide some exploratory ideas for design strategies in HRI.

The paper is structured as follows. In [Section 2](#), after a brief overview of the notion of joint action and its underlying mechanisms, we present some well-known challenges of social robotics for establishing interpersonal relations between humans and robots. We suggest that these challenges can be addressed through an emphasis on the communicative mechanisms present in joint action, including those found in human-human interactions. In [Section 3](#), we review different studies in developmental psychology, cognitive psychology, and philosophy of mind and language to demonstrate how such literature can help meet some of the challenges in HRI. We describe several mechanisms that may provide perspectives for social robotics in order to equip robots with robust communication capacities for joint action.

## 2. HRI and joint action: (some) current challenges

### 2.1. How can we define joint action?

An important number of social interactions and encounters are encompassed by the notion of *joint action*. Broadly considered, joint action is any form of social interaction whereby two agents or more coordinate their actions in order to pursue a joint goal. However, the notion of joint action has been subject to debate in philosophy and psychology. For instance, according to [Sebanz et al. \(2006\)](#), joint actions require the partners to coordinate “their actions in space and time to bring about a change in the environment” (p. 70); while other authors ([Carpenter, 2009](#); [Cohen & Levesque, 1991](#); [Fiebich & Gallagher, 2012](#); [Tomasello et al., 2005](#)) resist the idea that instances of mere coordination – e.g. two partners walking side by side – constitute a joint action, considering that it requires some necessary conditions like sharing goals and intentions.

Moreover, while the notion of joint action is used interchangeably with the notion of collaboration or cooperation for some authors ([Becchio et al., 2010](#); [Kobayashi et al., 2018](#)), other authors ([Amici & Bietti, 2015](#); [Chalmeau & Gallo, 1995](#)) establish a hierarchy of interactions

depending on the processes involved. According to [Amici and Bietti \(2015\)](#), for example, coordination is a fast low-level process of behavioral matching and interactional synchrony which could, but not necessarily, facilitate middle-level processes like cooperation (where some individuals bear certain costs to provide benefits to others) or high-level processes like joint action, which requires other resources like turn-taking and alignment of linguistic resources during dialogue.

### 2.2. What is necessary for joint action?

Leaving aside the debate on the concept of joint action, we aim to focus on the mechanisms that enable the consecution of joint actions. Three interrelated mechanisms appear to be key conditions for joint action: coordination, planning, and motivational alignment, each of them being supported by other processes. There has been an important deal of conceptual and empirical work investigating these processes ([Knoblich et al., 2011](#); [Pacherie, 2012](#); [Vesper et al., 2016](#)), from the sharing of a common ground to the anticipation of a partner’s actions by way of emergent coordination ([Curioni et al., 2017](#)).

Most of all, joint actions require individuals to anchor their plans in the actual situation and generate particular coordinated actions ([Knoblich et al., 2011](#); [Vesper et al., 2016](#)). This coordination can rely on different mechanisms, which are not necessarily intentional, including for example entrainment or rhythmic synchronization ([Harrison & Richardson, 2009](#); [Keller et al., 2014](#); [Richardson et al., 2007](#); [Schmidt & O’Brien, 1997](#)), perception-action matching ([Brass et al., 2001](#); [Brass & Heyes, 2005](#)), perception of joint affordances ([Ramenzoni et al., 2008](#)), emotion understanding ([Michael, 2011](#)), joint attention ([Emery, 2000](#); [Kourtit et al., 2014](#)), rational co-efficiency ([Török et al., 2019](#)) or action simulation ([Blakemore & Decety, 2001](#); [Sebanz & Knoblich, 2009](#)). Partners can thus execute so-called ‘coordination smoothers’ to facilitate coordination ([Vesper et al., 2010](#)); they can reduce the temporal variability of their actions ([Vesper et al., 2011](#)) or adapt their actions all the more so when they have an easier sub-task ([Skewes et al., 2015](#)) to improve coordination.

More specifically, intentional coordination – sometimes referred to as *planned coordination* ([Curioni et al., 2017](#)) – requires the partners: (i) to represent their own and others’ actions, as well as the consequences of these actions, (ii) to represent the hierarchy of sub-goals and sub-tasks of the plan, (iii) to generate predictions of their joint actions, and (iv) to monitor the progress toward the joint goal in order to possibly compensate or help others to achieve their contributions ([Pacherie, 2012](#)).

Indeed, joint action often involves planning several aspects, which involves the representation of the goal and the whole plan, and/or even the sequence of actions to be performed. The formation of these types of representations could rely on different mechanisms, which include high-level processes such as theory of mind, team reasoning, or verbal negotiation ([Bratman, 2014](#)) or more low-level processes as minimally representing the joint action goal and knowing that it will be achieved with others ([Michael & Székely, 2019](#); [Vesper et al., 2010](#)). An example of the mechanisms involved in planning a joint action is task co-representations, which allows individuals to represent the details of each other’s task. Several studies have demonstrated that people tend to represent others’ tasks even when it is pernicious for their own task performance ([Eskenazi et al., 2013](#); [Sebanz et al., 2003, 2005](#)). Such capacity appears early in development; for instance, 5-years-old children can incorporate their partner’s role into their own action plan

<sup>1</sup> The concerns raised by this otherwise necessary compartmentalisation of work not only come from the experience of some of us, but are also a recurring debate in social robotics (see, e.g., [Belhassein et al., 2020](#), [Menezes et al., 2014](#), [Seibt et al., 2020](#), [Young et al., 2011](#)).

during a joint activity, through the appearance of a joint Simon effect<sup>2</sup> (see Saby et al., 2014).

These representations allow individuals to generate predictions about the other's actions, which in turn facilitate the adjustment and coordination between the partners. Interestingly, individuals can also facilitate the others' and their own predictions by communicating relevant and reliable information for joint action. The objective is to make actions more transparent and predictable so that the decision-making on the interaction can be fluent and successful. An interesting example of these mechanisms is *sensorimotor communication*. Several studies suggest that actors exaggerate their movements or kinematic parameters to make their actions more understandable to partners (Sacheli et al., 2013; Vesper & Richardson, 2014). Besides such implicit communicative devices, participants in joint action often negotiate on the fly the sub-tasks, sub-goals, or ways to proceed regarding the collective task through different explicit exchanges (Clark, 1996). Finally, communicative mechanisms do not only improve coordination and prediction by providing relevant information about the specific course of action, but also by providing information that impacts the motivational forces of the partner; for instance, motivating the other to remain engaged in the joint task or fulfilling others' expectations (Heintz et al., 2015; Michael & Székely, 2018).

An important body of research in psychology and philosophy thus suggests that these communicative processes play a fundamental role in joint action. One of our goals is therefore to highlight the significance of these devices for social robotics, without which the flexibility and efficiency of human interactions to HRI would be impossible to transfer.

### 2.3. Joint action in HRI: attempts and challenges

#### 2.3.1. Attempts and success in social robotics

A highly influential view in social robotics holds that developing likable robots or whose appearance imitates human physical features can improve users' experience, and thus, interaction. In this view, social robotics must aim at constructing robots whose appearance and behavior may appeal to human positive emotions (e.g. curiosity or likability). Some examples of robots instantiating this strategy are Jibo (Jibo, Inc.) or Pepper (Softbank Robotics) whose rounded forms give them a pleasant and friendly appearance. Also, we can find humanoid robots like Erica or Geminoid, developed by Ishiguro and his colleagues (Glas et al., 2016; Nishio et al., 2007), which are as physically identical to humans as possible. This general strategy of taking advantage of human preferences, inclinations, or curiosity is not restricted to their physical features.

Several labs (see, for instance, Breazeal, 2002; Craig et al., 2010; Kishi et al., 2013; Oberman et al., 2007; Wendt et al., 2008; Wendt & Berg, 2009) are, for instance, equipping robots with different emotional expressions in order to prompt empathy or pro-social attitudes into the human agent, so robots "could potentially tap into the powerful social motivation system inherent in human life, which could lead to more enjoyable and longer-lasting human-robot interactions" (Oberman et al., 2007, p. 2195). For instance, Craig et al. (2010) have demonstrated that facial expressions of the BERT humanoid robots can elicit the same neuronal responses as human facial expressions in an emotion recognition task. Robotic emotional expressions can be implemented in very different ways, which are likely to operate as communicative cues. For instance, while many works involve facial expressions in anthropomorphic faces (Ahn et al., 2012; Kedzierski et al., 2013; Lütkebohle

et al., 2010), other implementations involve posture (Breazeal et al., 2007), body motion (Kishi et al., 2013) or pace patterns (Karg et al., 2010). Moreover, pro-social attitudes and emotional states can be elicited through eye gazes, gestures, or speed of movement. Ham et al. (2015) have demonstrated that human users find robots more persuasive when they orient their eyes and heads toward them, and Riek et al. (2010) have shown that equipping robots with cooperative gestures (beckon, give, shake hands) influence the human motivation for joint action. Finally, Wendt et al. (2008) have elicited different human emotional states (stress, boredom, surprise, and perplexity) using a robot arm in a LEGO building block by modifying speed and intervals.

Moreover, some of the robot's gestures or legible motions can also improve prediction in joint action, through kinematic signaling (Beetz et al., 2010; Dehais et al., 2011; Dragan et al., 2015; Holladay et al., 2014; Huang & Thomaz, 2010; Kruse et al., 2013; Lichtenthäler & Kirsch, 2013; May et al., 2015; Riek et al., 2010; Sisbot & Alami, 2012; Trujillo et al., 2019). For example, Holladay et al. (2014) proposed a mathematical model that generates pointing configurations making the target object easier to recognize by novel users. Moreover, Dragan et al. (2015) have demonstrated that legible motions planned to express the robot's intents lead to more fluent collaborations than motions planned to match the partner's expectations or than functional motions. These findings are confirmed by other studies showing that stereotypical motions, along with straight lines and additional gestures (see Lichtenthäler & Kirsch, 2013 for a review) are pivotal factors for legible robot behavior. In the same vein, different strategies have been explored in connection with anticipation of motions (Coovert et al., 2014; Szafir et al., 2015; Triebel et al., 2016). For instance, Khambhaita et al. (2016) equipped Spencer, a socially aware service robot, with the capacity for anticipating its next movement by looking at the target direction to improve social motion and avoid collisions, while Coovert et al. (2014) (see also Chadalavada et al., 2015) used projections of visual arrows and a simplified map to communicate the intended movements. Finally, Huang and Thomaz (2010) have demonstrated the impact of soliciting gestures and gazes for ensuring joint attention between humans and robots.

Certainly, communication is bidirectional, and thus requires not only producing signals and providing information but also understanding and interpreting the others' signals. In this sense, a substantial effort has been directed to equip robots with capacities for understanding different human social cues, including gazes, gestures or facial expressions (e.g. Alazrai & Lee, 2012; Benamara et al., 2019; Boucenna et al., 2014; Burger et al., 2012; Liu & Wang, 2018). For example, Droeschel et al. (2011) used two cameras and two laser range finders to detect human gazes or pointing gestures, while Anjum et al. (2014) used a Microsoft Kinect camera and Support Vector Machine to recognize eight activities (e.g., sit and drink or wave hello) with extreme accuracy. Lang et al. (2013) used a recognition mechanism able to localize faces and extract their features to attribute different emotional states (happiness, sadness, fear, etc.) in an object teaching scenario. Finally, Kulic and Croft (2007) used a classifier based on a Hidden Markov Model in order to estimate affective states from physiological data collected in HRI.

Now, how can these efforts be oriented to design more efficient social robots, able to engage in joint action? The aforementioned capabilities like recognizing human gestures and facial expressions, or producing eye gazes are oriented to reducing different types of uncertainties and thus fostering the readiness to interact, by providing different pieces of information that become common knowledge. However, despite the enormous advances regarding the equipment of robotic agents with socio-cognitive capacities, the attempts to establish a better mutual understanding between humans and robots have not always turned out to be successful. Recent studies in HRI have demonstrated that constructing social robots with social "ingredients" does not necessarily improve the interaction or the human experience with the robot (Giger et al., 2019; Sciutti et al., 2018).

<sup>2</sup> In psychology, the Simon effect describes the influence of a stimulus location (shown on a screen) in a two-choice reaction time task, leading to a better performance when stimulus and response are on the same side. The joint Simon effect refers to a social situation in which responses are divided between two participants, which results in a "Go/NoGo task". A spatial compatibility effect is also observed in this joint task.

### 2.3.2. Obstacles and challenges in social robotics

Empirical literature assessing the users' experience with robots has detected several elements that may undermine interactions between human and robotic agents, concerning human motivation. Different studies in HRI have suggested that humans are not always inclined to interact with robotic agents. An often-voiced problem is the well-known Uncanny Valley Effect (Wang et al., 2015 for a review). The uncanny valley effect refers to the phenomenon whereby humans experience a feeling of discomfort or revulsion when perceiving a machine or artifact that acts or looks like a human. The first reference to this effect appears in the work of Mori (1970), and has been observed not only in human adults but also in children (Yamamoto et al., 2009). Several studies have suggested that the uncanny valley is often associated with the robot appearance (Wang et al., 2015). However, other studies also suggest that the effect may be associated with certain non-verbal behaviors (Thepsonthorn et al., 2021) or, more generally, with any cue from which one may infer the existence of the robot's mind (Gray & Wegner, 2012), what seems to suggest that not only appearance but also implementing certain actions and skills in robots may have drawbacks for HRI. Experiencing negative feelings and discomfort when perceiving or interacting with a robot can interfere with the motivation of humans to engage in social interactions with robots. This could be especially important in contexts where the relation between the human and the robot must be tied, for instance for robotic companions who assume roles for elder care, teaching and childcare or therapy. In fact, as Hoffman (2020) has suggested, the uncanny valley effect can take a different form in this kind of contexts, namely, as "a resistance to accept the social roles and behaviors of a robot, especially in companionate and relational settings" (p. 535). This *Social Uncanniness* may undermine the motivation to interact with robots and, although it does not necessarily mean abandoning collective action, it can fuel negative aspects that could be detrimental to the HRI.

Other potential negative attitudes toward robots involve implicit biases and low levels of trust. First, in a series of experiments including implicit association tests, which measure the reaction times depending on the associations between a target (a robot or a human) and positive or negative attributes, Sanders et al. (2016) have found that participants exhibit implicit negative attitudes toward robots, even when they express explicit positive assessment toward them. Those experiments seem to demonstrate that people exhibit certain negative attitudes toward robots, or at least, less positive stances than the ones they direct toward humans. Second, our everyday interactions with humans require an important and appropriate level of trust. Several social transactions make us vulnerable to the other's actions, to the extent that the consequence of the transactions is based on the expectation that the other party will behave as he should. However, we have reasons to believe that several characteristics of HRI may influence this level of trust. For instance in a recent meta-analysis, Hancock et al. (2011) suggest that the level of trust in robots may be influenced by performance-based factors (reliability, false alarms, or rate-failures) and robot attributes (proximity, anthropomorphism, or personality). These factors can undermine the interaction in different ways. On the one hand, for instance, a robot with a pleasant personality or anthropomorphic features may induce a high level of trust, which may lead humans to generate unrealistic expectations regarding the robotic agent, over-rely on the latter, and then limit the monitoring of some aspects of the interaction. This "expectations gap", once faced with the real capabilities of the robotic agent (see also Kwon et al., 2016), may undermine the interaction itself, and thus deteriorate the motivation of the user to engage with the robots on subsequent occasions. On the other hand, an initially low level of trust may produce resistance to start interacting with the robot or provoke her to abandon the collaborative task, ending up in the disuse of the robot (Lewis et al., 2018). The lack of trust is especially problematic in work contexts where a major part of the main task or relevant sub-tasks must be carried out by a mechanical agent. Imagine, for example, what it would be like to work on an assembly line where blushing mechanical

arms are unreliable, or in commercial contexts where a blushing colleague does not carry out his tasks reliably and reliably.

Summing up, several features in HRI can produce aversive and negative attitudes or influence the level of trust, which can damage the motivation of the human to interact with robotic agents.

In addition to motivational aspects, the interaction with robots may also be impaired by prediction issues. The ideomotor theory, which postulates that we initiate our actions by predicting their effects in the environment (James, 1890; Shin et al., 2010; Stock & Stock, 2004), laid the first foundations for the idea that the mental representation of an action or its effects would activate the motor codes for that same action and thus cause a tendency to perform it (Sebanz & Knoblich, 2009). Similarly, the observation of an action performed by others activates our own motor system (Heyes, 2011). This phenomenon of motor resonance was first evidenced by studies demonstrating the existence of mirror neurons in monkeys (Gallese et al., 1996; Rizzolatti et al., 2001) and humans (Heyes, 2010; Kilner et al., 2009), activating both during the execution of a movement but also when they observed the same action generated by others. In the same way, the execution of an action can be hindered by the observation of another action by others through motor contagion (Bouquet et al., 2011). This activation of our own motor system during the planning, the mental representation or the observation of an action (Prinz, 1997) is notably explained by the theory of event coding (Hommel, 2009), which postulates that the representation of actions is done via networks called event files and corresponding to the different characteristics of the perceived effects of these same actions. Similarly, we would predict the effects of our actions before executing them through generative or integrative forward models (Pesquita et al., 2018; Wolpert et al., 2003). Importantly, the prediction of actions and their effects appears to be essential for successful joint action (Curioni et al., 2019; Sebanz et al., 2006) by among other things, enabling interpersonal coordination during motor interactions (Sacheli et al., 2021). Finally, a study by Vesper and Richardson (2014) showed that even without a direct perception of the other's action, partners involved in a joint action successfully predict the action of their co-agent and incorporate this prediction into their own action planning in order to complete the task successfully.

Several studies have investigated the influence of the agent's nature on the phenomenon of motor resonance, in particular in the case of a robotic agent. Indeed, humans must interpret the robot's motions and adapt their behavior to collaborate efficiently and safely; the study of motor resonance phenomenon in the context of HRI thus seems particularly relevant, especially to investigate the unconscious responses of humans to robotic agents (Sciutti et al., 2012). If some studies have shown no motor contagion during the observation of robotic motions (Kilner et al., 2003; Tai et al., 2004), therefore concluding that the mirror system is not activated by the observation of a robot's mechanical (non-biological) motion (Press et al., 2005), other imaging studies have nuanced this statement. In some cases, the mirror neuron system seems to be activated during the observation of mechanical motions performed by non-human agents (Gazzola et al., 2007 with an industrial robot arm; Oberman et al., 2007 with a mechanical five-fingered robot arm; Saby et al., 2011 with a hand puppet bear). By reproducing the study of Kilner et al. (2003) showing no motor resonance during the observation of a robotic agent, but this time using a humanoid robot, Atkeson et al. (2000) have highlighted a phenomenon of motor interference identical to the one appearing during human interactions, i.e. the observation of an incompatible action (carried out by the robot) disturbed the one the participants were executing. Motor contagion has also been found with a humanoid robot (Bisio et al., 2014), for movements whose trajectory was characterized by biological kinematics. In addition, the acquisition of bidirectional action-effect associations (action and its motor codes linked to the effects of this action in the environment) by observation, enabled by motor resonance mechanism, was found to be possible by the observation of a simple virtual agent presented via a computer screen (Belhassein et al., 2021). Finally, an EMG study has investigated how the

appearance and type of motion of an agent can influence the muscle activity of people observing or imitating videos (Hofree et al., 2015). More precisely, the authors demonstrated that motor simulation appears during observation and imitation of all types of agents (humans and robots with mechanical or biological appearance), but still with a stronger effect for human agents. In a nutshell, there is no consensus about whether or not motor resonance mechanisms can be successfully used to generate predictions regarding the robot's behavior.

Taken together, despite the considerable advances of social robotics for equipping robots with social skills, these empirical findings show that several factors can hinder the interaction between humans and robots, impacting in particular users' motivation and the possibility of prediction, which are two pivotal requirements for joint actions. Motivation is indeed fundamental for the partners to engage in the collaborative task, but also to remain engaged when more desirable individual options appear. Successful joint actions also require partners to efficiently coordinate in space and time, and adjust their behaviors to each other to reach the shared goal. Faced with the challenge of designing strategies to overcome or compensate for the impact of these elements in the context of joint action, social robotics has already tackled several issues. Developing this expertise requires focusing on specific capacities of the robotic agents, displayed in specific contexts. As a consequence, research goals in HRI are sometimes restricted to solving a very particular problem (e.g. avoiding collision, anticipating a movement, or using and tracking gazes) occurring in the course of joint action, which normally implies considering such capacity in isolation and under laboratory conditions that do not always represent real interaction conditions. However as we have seen above, joint action situations involve several types of uncertainties, including some that can emerge before the interaction itself.

Therefore, it appears necessary to go beyond the elements of design that serve an instrumental purpose regarding a given task and consider the whole robot as a social agent itself, who must engage proactively in contributing to a mutual understanding between the participants in the joint task. In our view, such objectives can be reached by focusing on the different communicative mechanisms involved in joint action, as alternative ways of sustaining motivation and re-establishing routes of prediction. In the following section, we review different findings and proposals in philosophy and psychology to provide exploratory ideas on some communicative strategies that could be implemented in HRI.

### 3. Addressing joint action challenges: some perspectives from psychology and philosophy

#### 3.1. Why focusing on communication?

An important part of human psychological devices involved in joint action is communicative, serving different purposes – e.g. negotiating, guiding, questioning (Austin, 1962; Clark, 1992; Sperber & Wilson, 1995) and leading to mobilize different types of information. This flexibility allows us to provide information about the relevant objects involved in a task, but also about the emotional or cognitive states of the participants. In the HRI context, this can help us to explore different routes or strategies to overcome the problems presented above that involve some degree of uncertainty in the different stages of joint action.

According to Michael and Pacherie (2015), participants can face three sources of uncertainty during joint action, which can overlap and influence each other. First, motivational uncertainty refers to the uncertainty of not knowing whether or not the partner is motivated to engage in the overall joint action, a particular goal, or sub-goal, or her degree of motivation. Second, instrumental uncertainty refers to the state of not knowing the other participant's instrumental beliefs on how to proceed, which roles to assume or when and where to act. Finally, common ground uncertainty emerges when instrumental beliefs and motivations are not mutually manifested. Thus, even if the participants share a goal or agree on how to proceed, they might not know that this is

the case. Any communicative act or strategy is directed to reduce common ground uncertainty, making mutually manifest a piece of information that can involve instrumental or motivational states, aspects of the environment, goals, or other relevant information for the consecution of the joint action. In a minimal sense, then, communicative strategies can be defined as overt stimuli –whether they are verbal or non-verbal– generated to activate, add up or update the common ground and knowledge related to a particular joint action.

Be that as it may, we contend that interacting with robots able to exhibit such communicative behavior can impact both motivation and prediction in the context of joint action. By making a piece of information (e.g. social roles) mutually manifest, communication can indeed influence humans' attitudes toward robots as potential partners, eliciting pro-social motivation, and also improve coordination and prediction by establishing several lines of understanding between robots and humans. Designing communicative strategies can thus have several functions likely to improve joint action. Studies in philosophy and psychology may offer some descriptions as potential guidelines for these strategies, including for example the notions of joint attention, commitments, or motivation alignment. The next section is devoted to present these potential guidelines, articulate how they facilitate the reduction of certain uncertainties –and thus, eliciting and facilitating coordination– and emphasize how they are directly relevant for HRI. Although we are aware that human-robot joint actions mainly involve non-verbal communicative strategies, we have chosen to emphasize verbal communication as well in some of the following sections, as it is a modality both powerful and spontaneous in humans. Many of the theories and notions that describe the dynamics of verbal communication (e.g., speech act theory) are also often used to characterize non-verbal behavior. Furthermore, as discussed above, there are reasons to believe that the uncanny valley effect caused by some non-verbal signals could potentially be counteracted by the verbal communication that accompany non-verbal behavior.

#### 3.2. Key processes for communication

##### 3.2.1. Joint attention and common ground

To perform a joint action, partners need a common goal. Indeed, joint action requires that individuals plan and perform their actions according to their predictions about the other's actions to reach this goal. *Joint attention* is a key feature for this purpose, as it allows the partners to establish and share a perceptual common ground, necessary to initiate the joint action but also for individuals already engaged in a joint action to coordinate successfully. Commonly defined as the ability to coordinate our attention to the same object of interest (e.g. Bakeman & Adamson, 1984), joint attention thus enables us to integrate others' attentional focus and therefore to experience the world together (Tomaseello, 1999). It is a key element of social cognition, playing a crucial role in « being and acting together » (Tomaseello & Carpenter, 2007).

The first studies of joint attention began with Bruner and collaborators in developmental psychology (Bruner, 1974; Scaife & Bruner, 1975). In a study with children aged 2–14 months old, the authors have shown that the ability to follow and share attention with others increases with age (Scaife & Bruner, 1975). Butterworth and Jarrett (1991) have determined three successive mechanisms involved in joint attention in the course of development: a first “ecological” mechanism by which children correctly follow the general direction of their parent's gaze, but cannot identify the target if there are two in their visual field; a “geometric” mechanism by 12 months that allows the child to follow precisely the other's gaze and identify the target's location; and finally a “representational” mechanism that appears between 12 and 18 months, thanks to which children can understand that their parent's gaze is directed toward a target outside of their visual field by forming a mental representation.

These episodes of joint attention, firstly massively initiated by the adult (Matthews et al., 2012), allow the child during his/her

development to progress from the simple sharing of attention by responding to the solicitations of the adult, to the direction of others' attention through initiation and continuity of joint attention (Carpenter et al., 1998; Mundy & Jarrold, 2010). Numerous studies have investigated the role of these joint attention processes on the development of other social and communicative abilities. In particular, children's ability to produce pointing gestures, illustrating their access to referential communication (e.g. Butterworth, 2003), has been shown, first, to rely on the ability to develop joint attention, and second, to contribute to its increasing complexity (Mundy, 2003; Vaughan Van Hecke et al., 2007). Joint attention has also been shown to predict the development of language skills: in a study with 6- to 18-month-old children, Morales et al. (1998) showed that the ability of 6-month-old children to follow the other's gaze predicts a better language vocabulary at 18 months old. Joint attention skills thus underpin the development of social cognition: they allow children, through an active role in the dynamics of social interaction (see Kidwell et al., 2007), to integrate the other as an attentional and intentional agent (Tomasello & Farrar, 1986). This founding role is particularly visible in atypical development, such as in children with autism who have difficulty communicating both verbally and non-verbally (Sigman et al., 1986) and who also show limited joint attention skills (see Bruinsma et al., 2004 for a review).

Several authors have added some complexity to this picture by going one step further in defining joint attention. While for some of them, two agents orienting their attention toward the same referent is a sufficient criterion to speak about joint attention (Butterworth, 1998; Butterworth & Jarrett, 1991), others have highlighted the need (1) to develop mutual knowledge of this coordinated attention, and (2) to represent the other agent's intentional states (Carpenter & Liebal, 2011). Tomasello (1995) for example defined joint attention as "people experiencing the same thing at the same time, and knowing together that they are doing this". Accordingly, joint attention can either be explained with basic learning mechanisms (lean joint attention) or as "the result of particular cognitive operations or second-order representational competencies" (Racine, 2011), i.e. what Tomasello and Carpenter (2007) called socio-cognitive abilities of shared intentionality (rich joint attention). The opposition between "lean" and "rich" views of joint attention may parallel the contrast between current research in the field of HRI, which tends to focus on "surface behaviors" (Kaplan & Hafner, 2006), like simultaneous looking or coordinated behaviors, and research in developmental psychology showing that joint attention encompasses a variety of socio-cognitive processes.

Examining the notions of *common ground* (or *common knowledge*) and *mutual recognition* may help reduce this gap, as they have also played a fundamental role in the definition of joint action in both philosophy and psychology (Alonso, 2009; Bratman, 1992, 1993; Clark, 1996; Cohen & Levesque, 1991; Lewis, 1969; Miller, 2001; Tollefsen, 2005; see also Blomberg, 2016 for criticism). First, Lewis (1969) claims that a proposition P is commonly known among two agents if the proposition is known by the two agents and both agents know that the other can draw the same conclusions from P that P can. In another famous formulation (see Schiffer, 1972; Thomas et al., 2014), common knowledge must be understood as the recursive belief in which S knows P, Y knows P, S knows that Y knows P, Y knows that S knows P, S knows that Y knows that S knows P, and so on. The subject does not necessarily represent the whole line of reasoning beforehand but should be able to infer it. Thus, we can assume that from the individual point of view, common knowledge or common ground is the information that one may reasonably assume that one and her partner know and they can also know or infer that the other knows. For our purpose, such information may include goals and sub-goals, intentions (see Bratman, 1992), ways to proceed, instrumental beliefs, facts on the environment, appropriate scripts and roles, and any other type of information necessary or relevant for the consecution of the joint action.

Second, the sharing of common ground is closely linked to what philosophers call *mutual recognition* (Brandom, 2007; Satne, 2014;

Scanlon, 1998), which allows individuals to identify and accept each other as social agents. Such mutual recognition requires each individual to identify the other as a partner for the available interaction, to generate expectations and anticipations depending on different physical and social features, information concerning previous interactions, or social structures like norms or conventions. Moreover, mutual recognition requires implicit or explicit confirmation that the other, as a social partner, somehow accepts the interaction. One of the safest ways to establish this recognition is to rely on communicative strategies.

Robots are regularly faced with failure situations because of common ground uncertainty or poor coordination during joint action. Therefore, joint attention appears to be an essential component for human-robot interactions and leads to significant advances in HRI, as shown by the numerous studies on this topic (see Admoni & Scassellati, 2017; Boucher et al., 2012; Huang & Thomaz, 2010; or Staudte & Crocker, 2009). Indeed, because joint attention allows sharing perceptual common ground, it leads to improved coordination, reduced uncertainties, and helps to perform better to reach the common goal. Similarly, it is necessary to establish mutual recognition and so to recognize the robot as a social partner in order to make human-robot interactions less ambiguous.

### 3.2.2. Communication for mutual recognition: examples of *recognitives* and *observatives*

The recognition of the other as a potential partner for joint action can be carried out by verbal and/or non-verbal communicative cues, which can be more or less explicit at different stages of the interaction. The inferential processes at play in such contexts have originally been explored in the frame of pragmatic theories, in particular through the notions of relevance (Sperber & Wilson, 1995) or Grice (1989). Interestingly, humans often establish communicative strategies to facilitate information exchange before the joint action itself, even in situations where social norms, conventions, or scripts are available to regulate our social interactions (Andrews, 2012; Fernández Castro & Heras-Escribano, 2020; Schank & Abelson, 1977; Zawidzki, 2013). For instance, as customers, we usually know how to interact with a waiter in a restaurant because the parties involved know some clear rules of etiquette, social norms and knowledge of how to proceed that regulate the interaction to achieve the joint goal of having a meal. However, even when these rules and norms exist, human interactions require signaling and communicating different types of information regarding the initiation, maintenance, or the exit of joint action, the acknowledgment of roles assignment, or specificities regarding preferences, goals, and sub-tasks. One can engage in communication employing so-called *recognitives* or *observatives*, speech acts whose main function is to call another person's attention upon herself, or other aspects of the context in order to make her aware that recognition is in place.

An example of *recognitives* is *vocatives*, like greetings that are precisely used to call a person upon herself. Vocatives can enable mutual recognition and facilitate role assignment in some contexts (e.g. "Welcome to our restaurant!" in the previous example). Moreover, vocatives are often followed by other speech acts like questions that can help to set the sub-tasks or goals of the joint action (e.g. "What can I do for you today?"). Another example of *recognitives* is acknowledgments, whose function is to make the other aware that you recognize or take on what they say (e.g. answering "thank you" to the vocative "welcome"). They allow individuals to acknowledge each other's recognition and to ensure that the fact joint action will take place is mutually shared.

The other types of speech act relevant for mutual recognition are *observatives*, which serve to identify a potential joint goal by directing the other's attention toward a specific object or event in the near environment. For instance, imagine two hunters searching for prey; when one calls the other "Hey, a deer!", they can start coordinating to capture the animal. Such speech acts can facilitate the recognition of the other as a potential partner for the joint action and then trigger the set of expectations and anticipations necessary to coordinate and perform the

action.

In a nutshell, humans have a whole set of speech acts at their disposal to facilitate the establishment of mutual recognition between partners and then the initiation of joint action. These analyses are further supported empirically by studies showing for example how adults (Brosnan et al., 2012) and 4-year-old children (Duguid et al., 2014) use verbal communication to achieve common knowledge and solve the problem of deciding whether or not to cooperate in some contexts.

Even though the verbal modality can notably modify the common ground between the speakers (see also Clark, 1992; Stalnaker, 1978, 2002 for linguistic inputs), it is far from being the only way to establish mutual recognition between individuals. We can find non-verbal modalities of communication analogous to recognitive or observatives. For instance, communication can stem from subtle cues like the mere reaction to the presence of the other with a frown movement or the search for eye contact. As Brinck and Balkenius (2018) argue, by making eye contact, one individual is attending to the other attending to the first, which can implicitly be regarded as a joint commitment to interact in most social contexts. Such analog to vocatives can be also introduced by other embodied strategies such as widening of the eyes, partially opening of the mouth or suddenly stilling of the limbs, all likely to reflect a possibility and/or search for confirmation to establish a joint action (Reddy & Morris, 2004, p. 658).

Likewise, acts of acknowledgements can be performed non-verbally as well: people often direct each other's attention toward external objects or events through non-verbal reference, whether it involves vocalizations, gestures, and/or gazing (Bard, 1992; Bates, 1979; Brinck, 2008; Leavens et al., 2004; Leavens et al., 2005). Non-verbal reference includes four essential actions: a *preparatory behavior* that draws the observer's attention to the sender, a *communicative-intent indicating behavior* to signal the sender's attempt to share attention and interact face-to-face with the observer; a *referential behavior*, to orient the other's attention in the direction of the target object or event; and an *essentially intentional behavior* that orients back the attention to oneself to make sure they understand the act (Brinck, 2008, p. 122–123).

Another way of establishing mutual recognition through non-verbal communication is to rely on ostensive acts, which can be regarded as non-verbal observatives. Notice for example how humans, from early infancy, are capable of predicting a course of action based on a target's hand movement (Koch & Stapel, 2019), sustained visual attention (Vaish & Woodward, 2010), or contextual constraints (Gergely et al., 2002). These capabilities are often used to build communicative strategies. For instance, Csibra and Gergely (2006) suggest that infants read some social signals –e.g. eye contact– as pedagogical signals: they indicate that what followed, for instance, a course of action, is an important piece of information. According to Csibra and Gergely (2006), these findings suggest that infants interpret eye contact as an “ostensive” act, understanding that the adult intends to communicate important information imminently. Moreover, parents often emit unconscious signals to reinforce such capacity, which generates a learning loop established by communication.

As a result, this type of communication aims here to orient the attention of the receiver to a particular object, but also to make her aware that the sender intends to share a particular purpose. Thus, in the context of joint action, this type of referential acts or even one of its components can play the same function as speech acts of acknowledgment: it can help initiate joint action and ensure that this possibility is common knowledge (Carpenter & Liebal, 2011; Chwe, 2001; Gómez, 1996; Thomas et al., 2014) or facilitate the recognition of a social affordance (Becchio et al., 2010; Krueger, 2011) –a property of an object or event to permit or forbid a social action. For instance, Sartori et al. (2009) investigated how the other's movements (e.g. extending the arm with an open hand toward an object) can perturb or influence the kinematics of a preplanned action. In this study, participants were asked to reach, pick and place an object; however, in 20% of the trials, the experimenter unexpectedly stretched out her arm and unfolds the hand

in a movement of request. A significant variation in the trajectory was reported and, in some cases, the participants even abandoned their plan to give the object to the experimenter. In contrast, no perturbation was found when the movement of the experimenter was arbitrary (i.e., showing no request features). These findings suggest that social signals can trigger social affordances and activate motor responses, which could facilitate the anticipation of behavior.

The establishment of mutual recognition is fundamental for the initiation of the joint action but also strongly influences its deployment. For instance, establishing mutual recognition facilitates the assignment of roles, which also determines the communicative strategies used during the execution of the action. The studies on the exaggeration of behavior mentioned in Section 2 illustrate this point: in Sacheli et al.'s (2013) experiments (see also Vesper & Richardson, 2014), for instance, two participants had to synchronously grasp an object in an imitative vs. complementary way, each by acting as a Leader or a Follower. The results showed that when acting as leaders, participants tend to give information to their partners about the action to be performed by accentuating some kinematic parameters and reducing the variability of movements, then increasing their predictability by the follower. Several research studies have further demonstrated that the assignment of Leader and Follower roles influenced participants in how they adapted their movements to improve interpersonal coordination in the joint action (Curioni et al., 2019). In an experiment with musicians, Goebel and Palmer (2009) have shown that the leader of a piano duets raised his fingers higher in the absence of auditory information, in order to communicate to his partner the movement timing. Those experiments indicate that signaling is fundamental during the performance of a task, even when the roles are somehow already assigned to the participants before engaging in joint action; it helps to coordinate interactions and optimize the joint action by minimizing uncertainties for each partner and thus increasing predictability regarding the sequence of actions (Pezzulo et al., 2013; Pezzulo & Dindo, 2011). These studies also illustrate that mutual recognition and the communicative strategies used to engage in joint action are fundamental, both to confirm the assigned roles in a given situation (thus allowing individuals to initiate the action), and to maintain joint action throughout the activity. In other words, to properly understand the importance of communication in joint action, we should take into account not only how a specific communicated information may influence the common ground at a given step, but also how it can impact subsequent steps of the action and the joint action as a whole (Vesper & Sevdalis, 2020).

Now, how is this communication for recognition relevant for HRI? To answer this question, we must emphasize that the aforementioned communicative strategies, as they facilitate the establishment of recognition, are important tools for reduction of uncertainties of different types (see Section 3.1). For instance, the use of recognitives or vocatives automatically reduces the motivational uncertainties by implicitly declaring that one is ready to interact. At the same time, the establishment of roles facilitates the reduction of common ground and instrumental uncertainties, to the extent that they can make explicit the rules and social structures at stake in the context and make mutually manifest certain general beliefs about how to proceed. The reduction of these two types of uncertainties is especially relevant in the context of HRI. On the one hand, we have seen how different factors like the novelty effect may produce certain reluctances to interact with the robot. These effects can be counterbalanced precisely by establishing recognition and expressing motivation to act. On the other hand, making environments less uncertain is especially relevant in HRI contexts as we know that, despite advances in social robotics, robots are still especially challenged with failures in unstructured contexts (Honig & Oron-Gilad, 2018). It is precisely the establishment of roles and the reduction of instrumental and common ground uncertainties that allows them to perform better in such environments. Moreover, the use of such strategies will become even more necessary when robots start to take on different roles in the same context, for example a robot that can take on the role of a co-



worker or a salesperson in the context of a shop.<sup>3</sup>

### 3.2.3. Commitment, emotion, and motivation

All the studies described above seem to demonstrate that communicative signals play a central role in the possibilities of prediction and coordination. By helping mutual recognition, these signals can also improve the motivation to initiate and perform the joint action, which has been theorized through the notion of commitments. When individuals search for eye contact with someone else, they are not only making the other aware of their availability to engage in a joint action, they are also implicitly declaring their commitment to interact and behave in accordance with this expectation. In a series of experiments, Siposova et al. (2018) presented 5 to 7-year-old children with a Stag Hunt coordination game, in which they could decide to play individually or to cooperate. In a first study, the authors compared two conditions with 5-year-olds: in the first condition, adults produced ostensive-communicative looks, with eyes widely open and raised eyebrows, while in the second condition, they produced non-communicative looks, i.e. shorter sequences without raised eyebrows. The authors' hypothesized that only communicative eye contact would permit mutual recognition and then encourage children to play cooperatively. In a second study, the authors investigated the reactions of 6 to 7-year-old children in case the partner did not cooperate. The results showed that children tend to cooperate more in the communicative eye contact condition and to protest more if the partner did not cooperate. These studies highlight the fact that communicative eye-contacts can be regarded as a form of commitment for cooperation and thus increase coordination. In other words, showing *communicative-intent indicating behavior* (see Section 3.2.2) can signal to the other that one is ready to initiate a particular joint action, which could provide motivation for that action.

Indeed, humans tend to interpret some social signals as implicit commitments or obligations to behave as the joint action demands. Michael and Pacherie (2015) have emphasized this connection between motivation, expectation, and commitments in joint actions. According to them, commitments play a fundamental role in a joint action by stabilizing expectations, which can reduce the uncertainties inherent in any interaction and provide reasons for cooperation. When people indicate that they are committed to perform a particular action, they are not only reducing motivational uncertainty –providing reliable cues about their action– but also providing reasons to cooperate on the basis that they are going to do their part, leading to the expectation that they expect the other to do the same.

The idea that we find other's expectations about our own behavior motivating and appealing by themselves is not new in philosophy and psychology. For instance, Lewis (1969) argues that we assume the existence of *presumptive reasons*: we are inclined to act to fulfill other's expectations when it is reasonable to have such expectations, i.e., under certain conditions. In a series of experiments, Heintz et al. (2015) have shown that people's pro-social preferences in dictator games, where a subject (the dictator) must decide whether or not to send some money he/she receives to another subject (the recipient), are sensitive to the other's expectations. In their studies, the dictator systematically exhibits more altruistic choices when he/she was told that the recipient expected his/her sending a particular amount of money. Similarly, Bonalumi et al. (2019) have tested the sensitivity to implicit signals about others' preferences in the context of joint action. In one of their studies, participants were presented with vignettes describing situations in which implicit commitments between characters were violated. For instance, a situation involves two work colleagues who always have coffee together in the same place, although they never explicitly agreed to do so. They keep having coffee for some time until one character doesn't show up. Several questions were then asked to participants to assess their

perception of the situation, in particular regarding the right of the character to demand an explanation to the other, and the assumed degree of negative emotions if no explanation was offered. The experimenters also introduced a temporal variable (e.g., how long had the coffee routine been going on?). They found that people tend to provide more negative emotions and negative normative judgments when the commitment was violated after a long routine period, compared to a more recent routine. This seems to demonstrate that the repetition of a joint action is perceived as an implicit cue to commitments, and that "people judge there to be an obligation to fulfill others' reasonable expectations even when these expectations have not been made explicit" (p. 685). In a nutshell, both implicit and explicit social signals regarding commitments enhance the motivation to engage in joint actions.

Moreover, Heintz et al. (2015); see also Sugden, 2000) postulate that human proclivity to engage in joint activities and fulfill others' expectations is caused by our aversion to disappointing others. In a similar perspective, Godman (2013); Godman et al., 2014) has argued that humans tend to engage in joint action due to a general psychological disposition to find affiliative stimuli rewarding (see also Salmela & Nagatsu, 2016). Interactions with others would then be intrinsically motivating and joint actions motivated not just by the desire to achieve an intended and shared outcome, but also by the desire to obtain this social reward. Evolutionary psychologists have theorized this idea of social reward through the notion of *reciprocity*: cooperating with others would be motivated by the assumption that one will receive benefits in return, benefits that can be direct or indirect and occur at different timescales (e.g., Romano & Balliet, 2017).

A more constrained hypothesis has been put forward by Fernández Castro and Pacherie (2021) who argue that, although humans give attentional priority to social cues and find them rewarding, their pro-social tendency lies in the *need to belong* (Baumeister & Leary, 1995; Over, 2016), i.e. the need for frequent, positively valenced interactions with other people within a framework of long-lasting concern for each other's welfare. In other words, joint actions are often motivated by the need to affiliate with others and form long-lasting bonds, to preserve and reinforce the bonds already forged.

Both implicit and explicit communicative cues can influence these motivational factors, reinforcing commitment and thus promoting joint action. Chartrand and Bargh (1999) have for instance demonstrated the motivational impact of the *chameleon effect*, which refers to the nonconscious mimicry of postures, gestures, and facial expressions when interacting with a partner. Individuals who were more likely to cooperate and who possessed more empathic dispositions were found to systematically exhibit the chameleon effect to a greater extent. These findings suggest that mimicry can be interpreted as a signal of the disposition for prosociality, therefore facilitating the initiation of joint action. Similarly, synchrony has been described as a rewarding social cue: Reddish et al. (2013) have found that moving together following the same rhythm also promotes prosocial behavior in economic games. The detection of these implicit communicative cues generally requires the cerebral activation of the amygdala, regarded as the integrative center for emotions and associated in particular with the reward system (Gamer et al., 2013).

More broadly, emotional expression may represent an important communicative strategy to facilitate joint action. Shared emotions, i.e., emotions whose expression has been directed to a partner and detected by the latter, have indeed been argued to play a pivotal role in motivating the initiation and the continuity of joint action (Michael, 2011). They can help to monitor performance and/or engagement: "a person's emotional expressions can transmit information about how she appraises her progress toward the goal of her own task, or the group's progress toward the global goal of a joint action" (p. 364). Thus, expressing emotions like excitement or enthusiasm toward the joint goal can provide the partner with the motivation to remain engaged in the joint action. Similarly, if one's performance during joint action provokes the expression of some distress or dislike in the partner, one could

<sup>3</sup> <https://anr.fr/Projet-ANR-20-IADJ-0006>

modify her behavior to avoid annoying or disappointing the other, therefore improving subsequent interactions. In short, emotional expressions can make the relevant information manifest for joint action.

The establishment and maintenance of commitments through communicative strategies, as well as the expression of certain shared emotions, is a potentially important tool for the design of social robotic agents. Insofar as they serve to make transparent and mutually manifest intentions to carry out a collective action, commitments and shared emotions have the capacity to reduce so-called motivational uncertainty, i.e., the lack of confidence or certainty that the other party will carry out the collective action. In the context of HRI, the communication of commitments and shared emotions could help us overcome two fundamental obstacles. First, during the initiation of the joint action, the use of these strategies could help neutralize users' possible reluctance to deal with robots, precisely because the display of a certain commitment from the robot might motivate the user to interact with it. Secondly, during the achievement of the action, the use of expressions and cues that help to maintain motivation is important in HRI given the opacity with which robots sometimes operate and the limited experience of users in dealing with robots. In fact, the use of emotional expressions to attract or maintain the users' attention have already been studied in some laboratories, for instance, using a sad face to inform users of a failure (Reyes et al. 2016).

#### 4. Concluding remarks

In the previous section, we have reviewed some recent literature in philosophy and psychology that highlights the triple function of communication in the context of joint action: from mutual recognition between individuals to the expression of commitment and social expectations, communicative cues can facilitate coordination, prediction, and motivation at different steps of the joint action. All the notions described above suggest that we can approach HRI through an integrated perspective of robotic communication in order to progress toward more robust social robots, able to engage in collaborative activities. Currently, while social robotics has dedicated a big effort to equip robots with specific capabilities, for example for recognizing human gestures or sending signals to make their actions more transparent, this expertise does not always translate into more fluid or more efficient interactions. This might be an effect of ever-increasing specialization that sometimes prevents researchers from considering the joint action as a whole, through the analysis of several communicative constraints and expectations both when initiating and when performing the action. For instance, the lack of communicative strategies dedicated to establishing mutual recognition before joint action may trigger previous unrealistic expectations of the human, which can undermine prediction during the whole process.

These difficulties can be overcome by adopting an integrated approach of HRI, i.e., an approach that provides us with a way to avoid compartmentalization (Belhassein et al., 2020; Menezes et al., 2014; Seibt et al., 2020; Young et al., 2011). Once we regard communication as a way to add up or modify the shared pool of relevant information in a given situation, every communicative device, rather than working in isolation to confront specific problems, can indirectly influence the information that impacts joint action as a whole. To establish mutual recognition between partners—a necessary prerequisite for initiating joint action—the robotic agent can for example restrict the type of shared information through scripts or social norms, making communicative signals more transparent. Verbal or gestural devices must therefore be oriented to make mutually manifest the different pieces of information that are relevant in the context of joint action regarding the partner's goals, plans, or beliefs. In other words, these devices can serve to constrain or add up information that facilitates the reduction of different types of uncertainties.

Such an integrative approach may also offer further flexibility: the design of communicative strategies can allow the robot to adapt to

different contexts, in which the roles assumed and the degree of common grounds shared with humans can vary. For example, the robot may have to interact differently with the human if the latter is a familiar partner or a stranger. Finally, this integration may provide some continuity to the interaction, which is particularly useful to determine the meanings of communicative signals. For example, recognizing a sign of frustration in a difficult collaborative context allows the agent to know that the partner has a problem. However, without a clear representation of the common ground to situate the specific action or sub-goal being carried out, the adjustment becomes more difficult. It can be even more challenging in situations where one agent detects a failure or an error that needs to be repaired, for instance, if the partner performs a wrong course of action. The continuity and stability of communication thus play a fundamental role not only in achieving objectives but also in motivating the agents to do so.

The present work aimed at suggesting some exploratory ideas to improve prediction and boost motivation in the context of joint action between humans and robots. The current state of the art in social robotics shows that some key ingredients are already available as a basis for developing communicative strategies, especially by focusing on human-robot contingencies (i.e., changes in an agent's behavior in direct response to a signal from another agent) as an indicator of the commitment in the interaction (Lee et al., 2011). A well-studied example is eye-gaze signaling, which has been proven to be an important source of understanding in HRI (Admoni & Scassellati, 2017; Kirchner et al., 2011; Moon et al., 2014; Staudte & Crocker, 2009). For instance, Breazeal et al. (2005) have demonstrated that equipping robots with subtle eye gazing signals—for instance, enabling Leonardo<sup>4</sup> to re-establish eye contact with the human when it finishes its turn, and then, communicating that it is ready to proceed to the next step in the task—improves the partners' understanding and allows them to quickly anticipate and address potential errors in the task. Similarly, human performance during a cooperative task with a robot improves significantly when people can follow the gaze of the robot (Boucher et al., 2012) or when the latter produces deictic gestures combined with directed gaze (Håring et al., 2012). Moreover, in ambiguous handover situations, people tend to comply with the direction of the robot's gaze (Admoni et al., 2014).

In addition, researchers are also seeking to design robots with perspective-taking skills to collaborate more efficiently (Trafton et al., 2005). During a human-robot joint activity, it can indeed be necessary for a robot to reason about the human mental states and knowledge of the situation to adapt its behavior (Scassellati, 2002) and make decisions (Görür et al., 2017). Devin and Alami (2016) developed a framework allowing this robotic Theory of Mind (ToM): the system they presented takes into account the human knowledge about the collaborative task in order to identify contingencies and provide her with the needed and only the needed information to reach the common goal.

The contribution of different disciplinary perspectives, in particular through discussions with psychology and philosophy, may therefore represent an important milestone to maximize current results in social robotics and improve the competence of the robot to collaborate with humans in different social environments. Focusing closely on communication between humans and robots during joint action could finally allow researchers to widen the application contexts (e.g. tasks, environments) of human-robot collaboration systems, possibly echoing Thomaz et al. (2016), for whom “implementing a collaborative planner in a complex realistic setting would be an apt grand challenge for the human-robot collaboration community”.

<sup>4</sup> Leonardo is a social and expressive robot created by the Personal Robots Group of the Massachusetts Institute of Technology, with communicative and manipulative functions, and a creature-like appearance.

## CRediT authorship contribution statement

**Kathleen Belhassein:** Writing – original draft, Conceptualization, Writing – review & editing. **Víctor Fernández-Castro:** Writing – original draft, Conceptualization, Writing – review & editing. **Amandine Mayima:** Writing – original draft, Writing – review & editing. **Aurélien Clodic:** Writing – review & editing, Supervision, Resources, Funding acquisition, Project administration. **Elisabeth Pacherie:** Resources, Funding acquisition, Project administration. **Michele Guidetti:** Writing – review & editing, Resources, Funding acquisition, Project administration. **Rachid Alami:** Writing – review & editing, Resources, Funding acquisition, Project administration. **Hélène Cochet:** Writing – original draft, Writing – review & editing, Supervision, Visualization.

## Declaration of competing interest

We hereby declare that there are none potential conflicts of interest that could bias the evaluation or results of our research.

## Acknowledgments

This research was supported by the French National Agency for Research grant No. ANR-16-CE33-0017 (JointAction4HRI project). Elisabeth Pacherie's work was also supported by grants ANR-17-EURE-0017 FrontCog and ANR-10-IDEX-0001-02 PSL. Víctor Fernández-Castro's work was also supported by the Juan de la Cierva-Incorporación grant (IJC2019-040199-I) and two projects of the Spanish Ministry of Science (PID2019-108870GB-I00, PID2019-109764RB-I00).

## References

- Admoni, H., Dragan, A., Srinivasa, S. S., & Scassellati, B. (2014). Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction - HRI '14* (pp. 49–56). <https://doi.org/10.1145/2559636.2559682>
- Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction*, 6(1), 25. <https://doi.org/10.5898/JHRI.6.1.Admoni>
- Ahn, H. S., Lee, D.-W., Choi, D., Lee, D.-Y., Hur, M., & Lee, H. (2012). Difference of efficiency in Human-Robot Interaction according to condition of experimental environment. In S. S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, & M.-A. Williams (Eds.), *Social Robotics* (Vol. Vol. 7621, p. 219–227). Springer Berlin Heidelberg. doi: [https://doi.org/10.1007/978-3-642-34103-8\\_22](https://doi.org/10.1007/978-3-642-34103-8_22)
- Alazrai, R., & Lee, C. S. G. (2012). Real-time emotion identification for socially intelligent robots. In *2012 IEEE international conference on robotics and automation* (pp. 4106–4111). <https://doi.org/10.1109/ICRA.2012.6224587>
- Alonso, F. M. (2009). Shared intention, reliance, and interpersonal obligations. *Ethics*, 119(3), 444–475. <https://doi.org/10.1086/599984>
- Amici, F., & Bietti, L. M. (2015). Coordination, collaboration and cooperation: Interdisciplinary perspectives. *Interaction Studies*, 16(3), vii–xii. <https://doi.org/10.1075/is.16.3.001int>
- Andrews, K. (2012). *Do apes read minds? In Toward a new folk psychology*. MIT Press.
- Anjum, M. L., Ahmad, O., Rosa, S., Yin, J., & Bona, B. (2014). Skeleton tracking based complex human activity recognition using kinect camera. In M. Beetz, B. Johnston, & M.-A. Williams (Eds.), *Social Robotics* (Vol. Vol. 8755, p. 23–33). Springer International Publishing. doi: [https://doi.org/10.1007/978-3-319-11973-1\\_3](https://doi.org/10.1007/978-3-319-11973-1_3)
- Atkeson, C. G., Hale, J. G., Pollick, F., Riley, M., Kotosaka, S., Schaul, S., ... Kawato, M. (2000). Using humanoid robots to study human behavior. *IEEE Intelligent Systems*, 15(4), 46–56. <https://doi.org/10.1109/5254.867912>
- Austin, J. (1962). *How to do things with words*. Clarendon Press.
- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 55(4), 1278. <https://doi.org/10.2307/1129997>
- Bard, K. A. (1992). Intentional behavior and intentional communication in young free-ranging orangutans. *Child Development*, 63(5), 1186. <https://doi.org/10.2307/1131526>
- Bates, E. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529. <https://doi.org/10.1037/0033-2909.117.3.497>
- Becchio, C., Sartori, L., & Castiello, U. (2010). Toward you: The social side of actions. *Current Directions in Psychological Science*, 19(3), 183–188. <https://doi.org/10.1177/0963721410370131>
- Beetz, M., Stulp, F., Esden-Tempski, P., Fedrizzi, A., Klank, U., Kresse, I., Maldonado, A., & Ruiz, F. (2010). Generality and legibility in mobile manipulation: Learning skills for routine tasks. *Autonomous Robots*, 28(1), 21–44. <https://doi.org/10.1007/s10514-009-9152-9>
- Belhassein, K., Fernández Castro, V., & Mayima, A. (2020). A horizontal approach to communication for human-robot joint action: Towards situated and sustainable robotics. In Seibt, Nørskov and Quick (Ed.), 335. *Culturally sustainable social robotics. Frontiers in artificial intelligence and applications* (pp. 204–214). IOS Press. <https://doi.org/10.3233/FAIA200916>
- Belhassein, K., Marshall, P. J., Badets, A., & Bouquet, C. A. (2021). Social learning of action-effect associations: Modulation of action control following observation of virtual action's effects. *Attention, Perception & Psychophysics*, 83, 484–496. <https://doi.org/10.3758/s13414-020-02157-1>
- Benamara, N. K., Val-Calvo, M., Álvarez-Sánchez, J. R., Díaz-Morcillo, A., Ferrández Vicente, J. M., Fernández-Jover, E., & Stambouli, T. B. (2019). Real-time emotional recognition for sociable robotics based on deep neural networks ensemble. In J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, J. Toledo Moreo, & H. Adeli (Eds.), *Understanding the brain function and emotions* (Vol. 11486, p. 171–180). Springer International Publishing. doi: [https://doi.org/10.1007/978-3-030-19591-5\\_18](https://doi.org/10.1007/978-3-030-19591-5_18)
- Bisio, A., Sciutti, A., Nori, F., Metta, G., Fadiga, L., Sandini, G., & Pozzo, T. (2014). Motor contagion during human-human and human-robot interaction. *PLoS One*, 9(8), Article e106172. <https://doi.org/10.1371/journal.pone.0106172>
- Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, 2, 561–567.
- Blomberg, O. (2016). Common knowledge and reductionism about shared agency. *Australasian Journal of Philosophy*, 94(2), 315–326. <https://doi.org/10.1080/00048402.2015.1055581>
- Bonalumi, F., Isella, M., & Michael, J. (2019). Cueing implicit commitment. *Review of Philosophy and Psychology*, 10(4), 669–688. <https://doi.org/10.1007/s13164-018-0425-0>
- Boucenna, S., Gaussier, P., Andry, P., & Hafemeister, L. (2014). A robot learns the facial expressions recognition and face/non-face discrimination through an imitation game. *International Journal of Social Robotics*, 6(4), 633–652. <https://doi.org/10.1007/s12369-014-0245-z>
- Boucher, J.-D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., ... Ventre-Dominey, J. (2012). I reach faster when i see you look: Gaze effects in human–human and human–robot face-to-face cooperation. *Frontiers in Neurobotics*, 6. <https://doi.org/10.3389/fnbot.2012.00003>
- Bouquet, C. A., Shipley, T. F., Capa, R. L., & Marshall, P. J. (2011). Motor contagion: Goal-directed actions are more contagious than non-goal-directed actions. *Experimental Psychology*, 58(1), 71–78. <https://doi.org/10.1027/1618-3169/a000069>
- Brage, A., Jean-Daubias, S., Loisel, E., & Basset, T. (2018). JOE: Le robot-compagnon des enfants asthmatiques. In *Conférence Nationale sur les Applications pratiques de l'Intelligence Artificielle, Nancy, France* (July). (hal-01811548).
- Brandon, R. B. (2007). The structure of desire and recognition: Self-consciousness and self-constitution. *Philosophy & Social Criticism*, 33(1), 127–150. <https://doi.org/10.1177/0191453707071389>
- Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, 106, 3–22.
- Brass, M., & Heyes, C. (2005). Imitation: Is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences*, 9(10), 489–495. <https://doi.org/10.1016/j.tics.2005.08.007>
- Bratman, M. E. (1992). Shared cooperative activity. *Philosophical Review*, 101(2), 327–341.
- Bratman, M. E. (1993). Shared intention. *Ethics*, 104(1), 97–113.
- Bratman, M. E. (2014). *Shared agency*. Oxford University Press.
- Breazeal, C. (2002). Regulation and entrainment in human-robot interaction. *International Journal of Robotics Research*, 21(10–11), 883–902. <https://doi.org/10.1177/0278364902021010096>
- Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 708–713). <https://doi.org/10.1109/IROS.2005.1545011>
- Breazeal, C., Wang, A., & Picard, R. (2007). Experiments with a robotic computer: Body, affect and cognition interactions. In *Proceeding of the ACM/IEEE International Conference on Human-Robot Interaction - HRI '07*, 153. <https://doi.org/10.1145/1228716.1228737>
- Brinck, I. (2008). The role of intersubjectivity in the development of intentional communication. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *Converging evidence in language and communication research* (Vol. Vol. 12, p. 115–140). John Benjamins Publishing Company. doi: <https://doi.org/10.1075/ceclr.12.08bri>
- Brinck, I., & Balkenius, C. (2018). Mutual recognition in human-robot interaction: A deflationary account. *Philosophy & Technology*, 33(1), 53–70. <https://doi.org/10.1007/s13347-018-0339-x>
- Brosnan, S. F., Wilson, B. J., & Beran, M. J. (2012). Old World monkeys are more similar to humans than New World monkeys when playing a coordination game. *Proceedings of the Royal Society B: Biological Sciences*, 279(1733), 1522–1530. <https://doi.org/10.1098/rspb.2011.1781>
- Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, 10(3), 169–175. <https://doi.org/10.1002/mrdd.20036>
- Bruner, J. S. (1974). From communication to language—A psychological perspective. *Cognition*, 3(3), 255–287. [https://doi.org/10.1016/0010-0277\(74\)90012-2](https://doi.org/10.1016/0010-0277(74)90012-2)
- Burger, B., Ferrané, I., Lerasle, F., & Infantes, G. (2012). Two-handed gesture recognition and fusion with speech to command a robot. *Autonomous Robots*, 32(2), 129–147. <https://doi.org/10.1007/s10514-011-9263-y>

- Butterworth, G. (1998). Origins of joint visual attention in infancy. *Monographs of the Society for Research in Child Development*, 63(4), 144–166. <https://doi.org/10.1111/j.1540-5834.1998.tb02782.x>
- Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 9–33). Lawrence Erlbaum Associates Publishers.
- Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, 9(1), 55–72. <https://doi.org/10.1111/j.2044-835X.1991.tb00862.x>
- 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>
- Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seeman (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (p. 159–181). MIT Press.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4), 1–143.
- Chadalavada, R. T., Andreasson, H., Krug, R., & Lilienthal, A. J. (2015). That's on my mind! Robot to human intention communication through on-board projection on shared floor space. In *2015 European Conference on Mobile Robots (ECMR)* (pp. 1–6). <https://doi.org/10.1109/ECMR.2015.7403771>
- Chalmeau, R., & Gallo, A. (1995). La coopération chez les primates. *L'Année Psychologique*, 95(1), 119–130. <https://doi.org/10.3406/psy.1995.28810>
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893–910. <https://doi.org/10.1037/0022-3514.76.6.893>
- Chwe, M. S.-Y. (2001). *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press.
- Clark, H. H. (1992). *Arenas of language use*. University of Chicago Press.
- Clark, H. H. (1996). *Using Language*. Cambridge University Press.
- Cohen, P. R., & Levesque, H. J. (1991). Teamwork. *Nous*, 25(4), 487–512.
- Coovert, M. D., Lee, T., Shinde, I., & Sun, Y. (2014). Spatial augmented reality as a method for a mobile robot to communicate intended movement. *Computers in Human Behavior*, 34, 241–248. <https://doi.org/10.1016/j.chb.2014.02.001>
- Craig, R., Vaidyanathan, R., James, C., & Melhuish, C. (2010). Assessment of human response to robot facial expressions through visual evoked potentials. In *2010 10th IEEE-RAS International Conference on Humanoid Robots* (pp. 647–652). <https://doi.org/10.1109/ICHR.2010.5686272>
- Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In *Vol. 21. Processes of change in brain and cognitive development. Attention and performance XXI* (pp. 249–274). Oxford University Press.
- Curioni, A., Knoblich, G., & Sebanz, N. (2017). Joint action in humans: A model for human-robot interactions. In A. Goswami, & P. Vadakkepat (Eds.), *Humanoid robotics: A reference* (pp. 1–19). Springer. [https://doi.org/10.1007/978-94-007-7194-9\\_126-1](https://doi.org/10.1007/978-94-007-7194-9_126-1)
- Curioni, A., Vesper, C., Knoblich, G., & Sebanz, N. (2019). Reciprocal information flow and role distribution support joint action coordination. *Cognition*, 187, 21–31. <https://doi.org/10.1016/j.cognition.2019.02.006>
- Dehais, F., Sisbot, E. A., Alami, R., & Causse, M. (2011). Physiological and subjective evaluation of a human-robot object hand-over task. *Applied Ergonomics*, 42(6), 785–791. <https://doi.org/10.1016/j.apergo.2010.12.005>
- Devin, S., & Alami, R. (2016). An implemented theory of mind to improve human-robot shared plans execution. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction - HRI '16* (pp. 319–326).
- Dragan, A. D., Bauman, S., Forlizzi, J., & Srinivasa, S. S. (2015). Effects of robot motion on human-robot collaboration. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15* (pp. 51–58). <https://doi.org/10.1145/2696454.2696473>
- Droeschel, D., Stueckler, J., Holz, D., & Behnke, S. (2011). Towards joint attention for a domestic service robot—Person awareness and gesture recognition using Time-of-Flight cameras. In *2011 IEEE International Conference on Robotics and Automation* (pp. 1205–1210). <https://doi.org/10.1109/ICRA.2011.5980067>
- Duguid, S., Wyman, E., Bullinger, A. F., Herfurth-Majstorovic, K., & Tomasello, M. (2014). Coordination strategies of chimpanzees and human children in a Stag Hunt game. *Proceedings of the Royal Society B: Biological Sciences*, 281(1796), 20141973. <https://doi.org/10.1098/rspb.2014.1973>
- Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6), 581–604. [https://doi.org/10.1016/S0149-7634\(00\)00025-7](https://doi.org/10.1016/S0149-7634(00)00025-7)
- Eskenazi, T., Doerrfeld, A., Logan, G. D., Knoblich, G., & Sebanz, N. (2013). 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>
- Fernández Castro, V., & Heras-Escribano, M. (2020). Social cognition: A normative approach. *Acta Analytica*, 35(1), 75–100. <https://doi.org/10.1007/s12136-019-00388-y>
- Fernández Castro, V., & Pacherie, E. (2021). Joint actions, commitments and the need to belong. *Synthese*, 198, 7597–7626. <https://doi.org/10.1007/s11229-020-02535-0>
- Fiebich, A., & Gallagher, S. (2012). Joint attention in joint action. *Philosophical Psychology*, 26(4), 571–587. <https://doi.org/10.1080/09515089.2012.690176>
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2), 593–609. <https://doi.org/10.1093/brain/119.2.593>
- Gamer, M., Schmitz, A. K., Tittgemeyer, M., & Schilbach, L. (2013). The human amygdala drives reflexive orienting towards facial features. *Current Biology*, 23(20), R917–R918. <https://doi.org/10.1016/j.cub.2013.09.008>
- Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, 35(4), 1674–1684. <https://doi.org/10.1016/j.neuroimage.2007.02.003>
- Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, 415(6873), 755. <https://doi.org/10.1038/415755a>
- Giger, J., Picarra, N., Alves-Oliveira, P., Oliveira, R., & Arriaga, P. (2019). Humanization of robots: Is it really such a good idea? *Human Behavior and Emerging Technologies*, 1(2), 111–123. <https://doi.org/10.1002/hbe2.147>
- Glas, D. F., Minato, T., Ishi, C. T., Kawahara, T., & Ishiguro, H. (2016). ERICA: The ERATO intelligent conversational Android. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 22–29. <https://doi.org/10.1109/ROMAN.2016.7745086>
- Godman, M. (2013). Why we do things together: The social motivation for joint action. *Philosophical Psychology*, 26(4), 588–603. <https://doi.org/10.1080/09515089.2012.670905>
- Godman, M., Nagatsu, M., & Salmela, M. (2014). The social motivation hypothesis for prosocial behavior. *Philosophy of the Social Sciences*, 44(5), 563–587. <https://doi.org/10.1177/0048393114530841>
- Goebel, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, 26(5), 427–438. <https://doi.org/10.1525/mp.2009.26.5.427>
- Gómez, J. C. (1996). Ostensive behavior in great apes: The role of eye contact. In A. E. Russon, K. A. Bard, & S. T. Parker (Eds.), *Reaching into thought: The minds of the great apes* (p. 131–151). Cambridge University Press.
- Görür, O. C., Rosman, B. S., Hoffman, G., & Albayrak, S. (2017). Toward integrating Theory of Mind into adaptive decision-making of social robots to understand human intention. In *Workshop The Role of Intentions in Human-Robot Interaction, at the 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*, Vienna, Austria.
- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125–130. <https://doi.org/10.1016/j.cognition.2012.06.007>
- Grice, H. P. (1989). *Studies in the way of words*. Harvard University Press.
- Ham, J., Cuijpers, R. H., & Cabibihan, J.-J. (2015). Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. *International Journal of Social Robotics*, 7(4), 479–487. <https://doi.org/10.1007/s12369-015-0280-4>
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(5), 517–527. <https://doi.org/10.1177/0018720811417254>
- Häring, M., Eichberg, J., & André, E. (2012). Studies on Grounding with Gaze and Pointing Gestures in Human-Robot-Interaction. In S. S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, & M.-A. Williams (Eds.), *Social robotics* (Vol. Vol. 7621, p. 378–387). Springer Berlin Heidelberg. doi:[https://doi.org/10.1007/978-3-642-34103-8\\_38](https://doi.org/10.1007/978-3-642-34103-8_38)
- Harrison, S. J., & Richardson, M. J. (2009). Horsing around: Spontaneous four-legged coordination. *Journal of Motor Behavior*, 41(6), 519–524. <https://doi.org/10.3200/35-08-014>
- Heintz, C., Celse, J., Giardini, F., & Max, S. (2015). Facing expectations: Those that we prefer to fulfil and those that we disregard. *Judgment & Decision Making*, 10(5), 442–455.
- Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience & Biobehavioral Reviews*, 34(4), 575–583. <https://doi.org/10.1016/j.neubiorev.2009.11.007>
- Heyes, C. (2011). Automatic imitation. *Psychological Bulletin*, 137(3), 463–483.
- Hofree, G., Urgen, B. A., Winkielman, P., & Saygin, A. P. (2015). Observation and imitation of actions performed by humans, androids, and robots: An EMG study. *Frontiers in Human Neuroscience*, 9. <https://doi.org/10.3389/fnhum.2015.00364>
- Hoffman, G. (2020). The Social Uncanniness of Robotic Companions. In Seibt, Nørskov, & Quick (Eds.), 335. *Culturally sustainable social robotics. Frontiers in artificial intelligence and applications* (pp. 204–214). IOS Press. <https://doi.org/10.3233/FAIA200953>
- Holladay, R. M., Dragan, A. D., & Srinivasa, S. S. (2014). Legible robot pointing. In *Proceedings of the 23rd IEEE international symposium on robot and human interactive communication* (pp. 217–223).
- Hommel, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research*, 73, 512–526. <https://doi.org/10.1007/s00426-009-0234-2>
- Honig, S., & Oron-Gilad, T. (2018). Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers of Psychology*, 9, 861. <https://doi.org/10.3389/fpsyg.2018.00861>
- Huang, C.-M., & Thomaz, A. L. (2010). Joint attention in human-robot interaction. In *AAAI fall symposium: Dialog with robots*.
- James, W. (1890). *The principles of psychology* (Vols. I, II). Cambridge, MA: Harvard University Press.
- Kanda, T., Shiomu, M., Miyashita, Z., Ishiguro, H., & Hagita, N. (2010). A communication robot in a shopping mall. *IEEE Transactions on Robotics*, 26(5), 897–913. <https://doi.org/10.1109/TRO.2010.2062550>
- Kaplan, F., & Hafner, V. V. (2006). The challenges of joint attention. *Interaction Studies*, 7(2), 135–169. <https://doi.org/10.1075/is.7.2.04kap>
- Karg, M., Kuhnlenz, K., & Buss, M. (2010). Recognition of affect based on gait patterns. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(4), 1050–1061. <https://doi.org/10.1109/TSMCB.2010.2044040>
- Kedzierski, J., Muszynski, R., Zoll, C., Oleksy, A., & Frontkiewicz, M. (2013). EMYS—Emotive head of a social robot. *International Journal of Social Robotics*, 5(2), 237–249. <https://doi.org/10.1007/s12369-013-0183-1>
- Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination.

- Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130394. <https://doi.org/10.1098/rstb.2013.0394>
- Khambhaita, H., Rios-Martinez, J., & Alami, R. (2016). Head-body motion coordination for human aware robot navigation. In *Proceedings of the 9th International Workshop on Human-Friendly Robotics (HFR2016)*.
- Kidwell, M., Zimmerman, & D. H. (2007). Joint attention as action. *Journal of Pragmatics*, 39, 592–611. <https://doi.org/10.1016/j.pragma.2006.07.012>
- Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience*, 29(32), 10153–10159. <https://doi.org/10.1523/JNEUROSCI.2668-09.2009>
- Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, 13(6), 522–525. [https://doi.org/10.1016/S0960-9822\(03\)00165-9](https://doi.org/10.1016/S0960-9822(03)00165-9)
- Kirchner, N., Alempijevic, A., & Dissanayake, G. (2011). Nonverbal robot-group interaction using an imitated gaze cue. In *Proceedings of the 6th International Conference on Human-Robot Interaction - HRI '11* (p. 497). <https://doi.org/10.1145/1957656.1957824>
- Kishi, T., Kojima, T., Endo, N., Destephe, M., Otani, T., Jamone, L., Kryczka, P., Trovato, G., Hashimoto, K., Cosentino, S., & Takanishi, A. (2013). Impression survey of the emotion expression humanoid robot with mental model based dynamic emotions. In *2013 IEEE International Conference on Robotics and Automation* (pp. 1663–1668). <https://doi.org/10.1109/ICRA.2013.6630793>
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action. In *Vol. 54. Psychology of learning and motivation* (pp. 59–101). Elsevier. <https://doi.org/10.1016/B978-0-12-385527-5.00003-6>
- Kobayashi, H., Yasuda, T., Igarashi, H., & Suzuki, S. (2018). Language use in joint action: The means of referring expressions. *International Journal of Social Robotics*. <https://doi.org/10.1007/s12369-017-0462-3>
- Koch, B., & Stapel, J. (2019). The role of head and hand movements for infants' predictions of others' actions. *Psychological Research*, 83(6), 1269–1280. <https://doi.org/10.1007/s00426-017-0939-6>
- Kourtis, D., Knoblich, G., Woźniak, M., & Sebanz, N. (2014). Attention allocation and task representation during joint action planning. *Journal of Cognitive Neuroscience*, 26(10), 2275–2286. <https://doi.org/10.1162/jocn.a.00634>
- Krueger, J. (2011). Extended cognition and the space of social interaction. *Consciousness and Cognition*, 20(3), 643–657. <https://doi.org/10.1016/j.concog.2010.09.022>
- Kruse, T., Pandey, A. K., Alami, R., & Kirsch, A. (2013). Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12), 1726–1743. <https://doi.org/10.1016/j.robot.2013.05.007>
- Kulic, D., & Croft, E. (2007). Physiological and subjective responses to articulated robot motion. *Robotica*, 25(01), 13. <https://doi.org/10.1017/S0263574706002955>
- Kwon, M., Jung, M., & Knepper, R. (2016). Human expectations of social robots. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 463–464).
- Lang, C., Wachsmuth, S., Hanheide, M., & Wersing, H. (2013). Facial communicative signal interpretation in human-robot interaction by discriminative video subsequence selection. In *2013 IEEE International Conference on Robotics and Automation* (pp. 170–177). <https://doi.org/10.1109/ICRA.2013.6630572>
- Leavens, D. A., Hopkins, W. D., & Thomas, R. K. (2004). Referential communication by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, 118(1), 48–57. <https://doi.org/10.1037/0735-7036.118.1.48>
- Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (Pan troglodytes). *Child Development*, 76(1), 291–306. <https://doi.org/10.1111/j.1467-8624.2005.00845.x>
- Lee, J., Kiser, J. F., Bobick, A. F., & Thomaz, A. L. (2011). Vision-based contingency detection. In *Proceedings of the 6th International Conference on Human-Robot Interaction - HRI '11* (p. 297). <https://doi.org/10.1145/1957656.1957780>
- Lewis, D. K. (1969). *Convention: A philosophical study*. Harvard University Press.
- Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. In *Foundations of trusted autonomy* (pp. 135–159). Springer.
- Lichtenhäger, C., & Kirsch, A. (2013). Towards legible robot navigation—How to increase the intend expressiveness of robot navigation behavior. In *International Conference on Social Robotics - Workshop Embodied Communication of Goals and Intentions*. <https://hal.archives-ouvertes.fr/hal-01684307>
- Liu, H., & Wang, L. (2018). Gesture recognition for human-robot collaboration: A review. *International Journal of Industrial Ergonomics*, 68, 355–367. <https://doi.org/10.1016/j.ergon.2017.02.004>
- Lütkebohle, I., Hegel, F., Schulz, S., Hackel, M., Wrede, B., Wachsmuth, S., & Sagerer, G. (2010). The bielefeld anthropomorphic robot head “Flobi”. In *2010 IEEE international conference on robotics and automation* (pp. 3384–3391). <https://doi.org/10.1109/ROBOT.2010.5509173>
- Matthews, D., Behne, T., Lieven, E., & Tomasello, M. (2012). Origins of the human pointing gesture: A training study. *Developmental Science*, 15(6), 817–829. <https://doi.org/10.1111/j.1467-7687.2012.01181.x>
- May, A. D., Dondrup, C., & Hanheide, M. (2015). Show me your moves!. In *Conveying navigation intention of a mobile robot to humans* (pp. 1–6). <https://doi.org/10.1109/ECMR.2015.7324049>
- McGlynn, S., Snook, B., Kemple, S., Mitzner, T. L., & Rogers, W. A. (2014). Therapeutic robots for older adults: Investigating the potential of paro. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction - HRI '14* (pp. 246–247). <https://doi.org/10.1145/2559636.2559846>
- Menezes, P., Quintas, J., & Miranda, D. J. (2014). The role of context information in human-robot interaction. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication, Workshop on Interactive Robots for aging and/or impaired people*.
- Michael, J. (2011). Shared emotions and joint action. *Review of Philosophy and Psychology*, 2(2), 355–373. <https://doi.org/10.1007/s13164-011-0055-2>
- Michael, J., & Pacherie, E. (2015). On commitments and other uncertainty reduction tools in joint action. *Journal of Social Ontology*, 1(1), 89–120. <https://doi.org/10.1515/jso-2014-0021>
- Michael, J., & Székely, M. (2018). The developmental origins of commitment. *Journal of Social Philosophy*, 49(1), 106–123. <https://doi.org/10.1111/josp.12220>
- Michael, J., & Székely, M. (2019). Goal slippage: A mechanism for spontaneous instrumental helping in infancy? *Topoi*. <https://doi.org/10.1007/s11245-017-9485-5>
- Miller, S. (2001). *Social action: A teleological account*. Cambridge University Press.
- Moon, A., Troniak, D. M., Gleeson, B., Pan, M. K. X. J., Zeng, M., Blumer, B. A., ... Croft, E. A. (2014). Meet me where I'm gazing: How shared attention gaze affects human-robot handover timing. In *Proceedings of the Ninth ACM/IEEE International Conference on Human-Robot Interaction (HRI'14)* (pp. 334–341). <https://doi.org/10.1145/2559636.2559656>
- Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development*, 21(2), 373–377. [https://doi.org/10.1016/S0163-6383\(98\)90014-5](https://doi.org/10.1016/S0163-6383(98)90014-5)
- Mori, M. (1970). Bukimi no tani (The uncanny valley). *Energy*, 7, 33–35.
- Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology and Psychiatry*, 44(6), 793–809. <https://doi.org/10.1111/1469-7610.00165>
- Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks*, 23(8–9), 985–997. <https://doi.org/10.1016/j.neunet.2010.08.009>
- Nishio, S., Ishiguro, H., & Hagit, N. (2007). Geminoid: Teleoperated android of an existing person. In A. C. de Pina Filho (Ed.), *Humanoid robots: New developments*. I-Tech Education and Publishing. doi:<https://doi.org/10.5772/4876>
- Oberman, L. M., McCleery, J. P., Ramachandran, V. S., & Pineda, J. A. (2007). EEG evidence for mirror neuron activity during the observation of human and robot actions: Toward an analysis of the human qualities of interactive robots. *Neurocomputing*, 70(13–15), 2194–2203. <https://doi.org/10.1016/j.neucom.2006.02.024>
- Over, H. (2016). The origins of belonging: Social motivation in infants and young children. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1686). <https://doi.org/10.1098/rstb.2015.0072>
- Pacherie, E. (2012). The phenomenology of joint action: Self-agency versus joint agency. In A. Seeman (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (p. 343–389). MIT Press.
- 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/s13423-017-1393-6>
- Pezzulo, G., & Dindo, H. (2011). What should I do next? Using shared representations to solve interaction problems. *Experimental Brain Research*, 211(3–4), 613–630. <https://doi.org/10.1007/s00221-011-2712-1>
- Pezzulo, G., Donnarumma, F., & Dindo, H. (2013). Human sensorimotor communication: A theory of signaling in online social interactions. *PLoS One*, 8(11), Article e79876. <https://doi.org/10.1371/journal.pone.0079876>
- Press, C., Bird, G., Flach, R., & Heyes, C. (2005). Robotic movement elicits automatic imitation. *Cognitive Brain Research*, 25(3), 632–640. <https://doi.org/10.1016/j.cogbrainres.2005.08.020>
- Prinz, W. (1997). *Perception and action planning*. *European Journal of Cognitive Psychology*, 9(2), 129–154.
- Racine, T. A. (2011). Getting beyond rich and lean views of joint attention. In A. Seeman (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (p. 21–42). MIT Press.
- Ramenzoni, V. C., Riley, M. A., Shockley, K., & Davis, T. (2008). Carrying the height of the world on your ankles: Encumbering observers reduces estimates of how high an actor can jump. *Quarterly Journal of Experimental Psychology*, 61(10), 1487–1495. <https://doi.org/10.1080/17470210802100073>
- Reddish, P., Fischer, R., & Bulbulia, J. (2013). Let's dance together: Synchrony, shared intentionality and cooperation. *PLoS One*, 8(8), Article e71182. <https://doi.org/10.1371/journal.pone.0071182>
- Reddy, V., & Morris, P. (2004). Participants don't need theories: Knowing minds in engagement. *Theory & Psychology*, 14(5), 647–665. <https://doi.org/10.1177/0959354304046177>
- Reyes, M., Meza, I., & Pineda, L. A. (2016). The positive effect of negative feedback in hri using a facial expression robot. In J. Koh, B. Dunstan, D. Silvera-Tawil, & M. Velonaki (Eds.), *vol 9549. Cultural robotics. CR 2015. Lecture notes in computer science*. Cham: Springer. [https://doi.org/10.1007/978-3-319-42945-8\\_4](https://doi.org/10.1007/978-3-319-42945-8_4)
- 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>
- Riek, L. D., Rabinowitch, T.-C., Bremner, P., Pipe, A. G., Fraser, M., & Robinson, P. (2010). Cooperative gestures: Effective signaling for humanoid robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 61–68). <https://doi.org/10.1109/HRI.2010.5453266>
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2, 661–670.
- Romano, A., & Balliet, D. (2017). Reciprocity outperforms conformity to promote cooperation. *Psychological Science*, 28, 1490–1502.

- Saby, J. N., Bouquet, C. A., & Marshall, P. J. (2014). Young children co-represent a partner's task: Evidence for a joint Simon effect in five-year-olds. *Cognitive Development*, 32, 38–45. <https://doi.org/10.1016/j.cogdev.2014.08.001>
- Saby, J. N., Marshall, P. J., Smythe, R., Bouquet, C. A., & Comalli, C. E. (2011). An investigation of the determinants of motor contagion in preschool children. *Acta Psychologica*, 138(1), 231–236. <https://doi.org/10.1016/j.actpsy.2011.06.008>
- Sacheli, L. M., Arcangeli, E., Carioti, D., Butterfill, S., & Berlinger, M. (2021). 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*. <https://doi.org/10.1177/17470218211050198>
- Sacheli, L. M., Tidoni, E., Pavone, E. F., Aglioti, S. M., & Candidi, M. (2013). Kinematics fingerprints of leader and follower role-taking during cooperative joint actions. *Experimental Brain Research*, 226(4), 473–486. <https://doi.org/10.1007/s00221-013-3459-7>
- Salmela, M., & Nagatsu, M. (2016). Collective emotions and joint action. *Journal of Social Ontology*, 2(1). <https://doi.org/10.1515/jso-2015-0020>
- Sanders, T. L., Schafer, K. E., Volante, W., Reardon, A., & Hancock, P. A. (2016). Implicit attitudes toward robots. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1), 1746–1749. <https://doi.org/10.1177/1541931213601400>
- Sandoval, E. B., Mubin, O., & Obaid, M. (2014). Human robot interaction and fiction: A contradiction. In M. Beetz, B. Johnston, & M. A. Williams (Eds.), *Proceedings of the international conference on social robotics (ICSR 2014)* (pp. 54–63). Cham: Springer. [https://doi.org/10.1007/978-3-319-11973-1\\_6](https://doi.org/10.1007/978-3-319-11973-1_6)
- Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., & Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *2011 Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI'11)* (pp. 305–311).
- 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), 1490–1500. <https://doi.org/10.1037/a0015777>
- Satne, G. (2014). What binds us together: Normativity and the second person. *Philosophical Topics*, 42(1), 43–61.
- Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, 253(5489), 265–266. <https://doi.org/10.1038/253265a0>
- Scanlon, T. (1998). *What we owe to each other*. The Belknap Press of Harvard University Press.
- Scassellati, B. (2002). Theory of mind for a humanoid robot. *Autonomous Robots*, 12(1), 13–24.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum.
- Schiffer, S. (1972). *Meaning*. Oxford University Press.
- Schmidt, R. C., & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, 9(3), 189–206. [https://doi.org/10.1207/s15326969eco0903\\_2](https://doi.org/10.1207/s15326969eco0903_2)
- Sciutti, A., Bisio, A., Nori, F., Metta, G., Fadiga, L., Pozzo, T., & Sandini, G. (2012). Measuring human-robot interaction through motor resonance. *International Journal of Social Robotics*, 4(3), 223–234. <https://doi.org/10.1007/s12369-012-0143-1>
- Sciutti, A., Mara, M., Tagliascio, V., & Sandini, G. (2018). Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technology and Society Magazine*, 37(1), 22–29. <https://doi.org/10.1109/MTS.2018.2795095>
- 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. (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>
- Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, 88(3), B11–B21. [https://doi.org/10.1016/S0010-0277\(03\)00043-X](https://doi.org/10.1016/S0010-0277(03)00043-X)
- 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>
- Seibt, J., Flensburg Damholdt, M., & Vestergaard, C. (2020). Integrative social robotics, value-driven design, and transdisciplinarity. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 21(1), 111–144. <https://doi.org/10.1075/is.18061.sei>
- Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, 136(6), 943–974.
- Sigman, M., Mundy, P., Sherman, T., & Ungerer, J. (1986). Social interactions of autistic, mentally retarded and normal children and their caregivers. *Journal of Child Psychology and Psychiatry*, 27(5), 647–656. <https://doi.org/10.1111/j.1469-7610.1986.tb00189.x>
- Siposova, B., Tomasello, M., & Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. *Cognition*, 179, 192–201. <https://doi.org/10.1016/j.cognition.2018.06.010>
- Sisbot, E. A., & Alami, R. (2012). A human-aware manipulation planner. *IEEE Transactions on Robotics*, 28(5), 1045–1057. <https://doi.org/10.1109/TRO.2012.2196303>
- Skewes, J. C., Skewes, L., Michael, J., & Konvalinka, I. (2015). Synchronised and complementary coordination mechanisms in an asymmetric joint aiming task. *Experimental Brain Research*, 233(2), 551–565. <https://doi.org/10.1007/s00221-014-4135-2>
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Blackwell Publishers.
- Stalnaker, R. (1978). Assertion. *Syntax and Semantics*, 9, 315–332.
- Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, 25(5/6), 701–721. <https://doi.org/10.1023/A:1020867916902>
- Staudte, M., & Crocker, M. W. (2009). Visual attention in spoken human-robot interaction. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction - HRI '09* (p. 77). <https://doi.org/10.1145/1514095.1514111>
- Stock, A., & Stock, C. (2004). A short history of ideomotor action. *Psychological Research*, 68(2–3), 176–188.
- Sugden, R. (2000). The motivating power of expectations. In J. Nida-Rümelin & W. Spohn (Eds.), *Rationality, Rules, and Structure* (p. 103–129). Springer Netherlands. doi:[https://doi.org/10.1007/978-94-015-9616-9\\_7](https://doi.org/10.1007/978-94-015-9616-9_7)
- Szafir, D., Mutlu, B., & Fong, T. (2015). Communicating directionality in flying robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15* (pp. 19–26). <https://doi.org/10.1145/2696454.2696475>
- Tai, Y. F., Scherfler, C., Brooks, D. J., Sawamoto, N., & Castiello, U. (2004). The human premotor cortex is « mirror » only for biological actions. *Current Biology*, 14(2), 117–120. <https://doi.org/10.1016/j.cub.2004.01.005>
- Thepsonthorn, C., Ogawa, K., & Miyake, Y. (2021). The exploration of the uncanny valley from the viewpoint of the robot's nonverbal behaviour. *International Journal of Social Robotics*, 13, 1443–1455. <https://doi.org/10.1007/s12369-020-00726-w>
- Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, 107(4), 657–676. <https://doi.org/10.1037/a0037037>
- Thomaz, A., Hoffman, G., & Cakmak, M. (2016). Computational human-robot interaction. *Foundations and Trends in Robotics*, 4(2–3), 104–223. <https://doi.org/10.1561/23000000049>
- Tollefsen, D. (2005). Let's pretend!: Children and joint action. *Philosophy of the Social Sciences*, 35(1), 75–97. <https://doi.org/10.1177/0048393104271925>
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (p. 103–130). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.
- Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, 10(1), 121–125. <https://doi.org/10.1111/j.1467-6872.2007.00573.x>
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(05), 675–691.
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57(6), 1454. <https://doi.org/10.2307/1130423>
- Török, G., Pomiechowska, B., Csibra, G., & Sebanz, N. (2019). Rationality in joint action: Maximizing efficiency in coordination. *Psychological Science*, 30(6), 930–941. <https://doi.org/10.1177/0956797619842550>
- Trafton, J. G., Schultz, A. C., Bugajska, M., & Mintz, F. (2005). Perspective-taking with robots: Experiments and models. In , 580-584. *IEEE International Workshop on Robot and Human Interactive Communication*. <https://doi.org/10.1109/ROMAN.2005.1513842>
- Triebel, R., Arras, K., Alami, R., Beyer, L., Breuers, S., Chatila, R., Chetouani, M., Cremers, D., Evers, V., Fiore, M., Hung, H., Ramirez, O. A. I., Joesse, M., Khambhaita, H., Kucner, T., Leibe, B., Lilienthal, A. J., Linder, T., Lohse, M., ... Zhang, L. (2016). SPENCER: A socially aware service robot for passenger guidance and help in busy airports. In D. S. Wettergreen & T. D. Barfoot (Eds.), *Field and service robotics: Results of the 10th international conference (Vol. 11)*, 613, p. 607–622. Springer International Publishing. doi:[https://doi.org/10.1007/978-3-319-27702-8\\_40](https://doi.org/10.1007/978-3-319-27702-8_40)
- Trujillo, J. P., Vaitonyte, J., Simanova, I., & Özyürek, A. (2019). Toward the markerless and automatic analysis of kinematic features: A toolkit for gesture and movement research. *Behavior Research Methods*, 51(2), 769–777. <https://doi.org/10.3758/s13428-018-1086-8>
- Vaish, A., & Woodward, A. (2010). Infants use attention but not emotions to predict others' actions. *Infant Behavior and Development*, 33(1), 79–87. <https://doi.org/10.1016/j.infbeh.2009.11.003>
- Vandemeulebroucke, T., de Casterlé, B. D., & Gastmans, C. (2018). How do older adults experience and perceive socially assistive robots in aged care: A systematic review of qualitative evidence. *Aging & Mental Health*, 22(2), 149–167. <https://doi.org/10.1080/13607863.2017.1286455>
- Vaughan Van Hecke, A., Mundy, P. C., Acra, C. F., Block, J. J., Delgado, C. E. F., Parlade, M. V., ... Pomares, Y. B. (2007). Infant joint attention, temperament, and social competence in preschool children. *Child Development*, 78(1), 53–69. <https://doi.org/10.1111/j.1467-8624.2007.00985.x>
- Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, 23(8–9), 998–1003. <https://doi.org/10.1016/j.neunet.2010.06.002>
- Vesper, C., & Richardson, M. J. (2014). Strategic communication and behavioral coupling in asymmetric joint action. *Experimental Brain Research*, 232(9), 2945–2956. <https://doi.org/10.1007/s00221-014-3982-1>
- Vesper, C., Schmitz, L., Safra, L., Sebanz, N., & Knoblich, G. (2016). The role of shared visual information for joint action coordination. *Cognition*, 153, 118–123. <https://doi.org/10.1016/j.cognition.2016.05.002>
- Vesper, C., & Sevdalis, V. (2020). Informing, coordinating, and performing: A Perspective on functions of sensorimotor communication. *Frontiers in Human Neuroscience*, 14. <https://doi.org/10.3389/fnhum.2020.00168>
- Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, 211(3–4), 517–530. <https://doi.org/10.1007/s00221-011-2706-z>

- Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, 19(4), 393–407. <https://doi.org/10.1037/gpr0000056>
- Wendt, C., Popp, M., Karg, M., & Kuhlentz, K. (2008). Physiology and HRI: Recognition of over- and underchallenge. In *RO-MAN 2008 - The 17th IEEE international symposium on robot and human interactive communication* (pp. 448–452). <https://doi.org/10.1109/ROMAN.2008.4600707>
- Wendt, C. S., & Berg, G. (2009). Nonverbal humor as a new dimension of HRI. In *RO-MAN 2009 - The 18th IEEE international symposium on robot and human interactive communication* (pp. 183–188). <https://doi.org/10.1109/ROMAN.2009.5326230>
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B*, 358, 593–602. <https://doi.org/10.1098/rstb.2002.1238>
- Yamamoto, K., Tanaka, S., Kobayashi, H., Kozima, H., & Hashiya, K. (2009). A non-humanoid robot in the “Uncanny Valley”: Experimental analysis of the reaction to behavioral contingency in 2–3 year old children. *PLoS One*, 4(9), Article e6974. <https://doi.org/10.1371/journal.pone.0006974>
- Young, J. E., Sung, J., Voids, A., Sharlin, E., Igarashi, T., Christensen, H. I., & Grinter, R. E. (2011). Evaluating human-robot interaction: Focusing on the holistic interaction experience. *International Journal of Social Robotics*, 3(1), 53–67. <https://doi.org/10.1007/s12369-010-0081-8>
- Zawadzki, T. (2013). *Mindshaping: A new framework for understanding human social cognition*. MIT Press.