Towards integrating joint action research: Developmental and evolutionary perspectives on co-representation

F.M. Miss\textsuperscript{a,\textdagger}, J.E.C. Adriaense\textsuperscript{a}, J.M. Burkart\textsuperscript{a,\textdagger,*}

\textsuperscript{a} Department of Anthropology, University of Zurich, Switzerland
\textsuperscript{b} Center for the Interdisciplinary Study of Language Evolution ISLE, University of Zurich, Switzerland

A R T I C L E   I N F O

Keywords:
Joint action
Primates
Co-representation
Evolution of cooperation
Social cognition
Ontogeny
Self-other integration and distinction
Cooperative breeding
Comparative research

A B S T R A C T

Joint action has increasingly become a key topic to understand the emergence of the human mind. The phenomenon is closely linked to several theoretical concepts, such as shared intentionality, which are difficult to operationalize empirically. We therefore employ a paradigm-driven, bottom-up approach, and as such discuss co-representing the partner’s and one’s own actions as key mechanism for joint action. After embedding co-representation in the broader landscape of related theoretical concepts, we review neurobiological, ontogenetic, and phylogenetic studies, with a focus on whether co-representation and its flexible deployment should be construed as a low- or high-level cognitive process. The empirical findings convergently suggest that co-representation does not require strong inhibitory skills or mentalistic understanding and occurs automatically. Moreover, more cooperative species are better at flexibly suppressing co-representation when required for cooperation success, and frequently rely on cooperation markers, such as mutual gaze. We thus contribute to closing the current gap between theoretical concepts related to joint action research and their empirical investigation, and end by highlighting additional approaches for doing so.

1. Introduction

Joint action is extensively studied in fields as diverse as developmental and comparative psychology, the social neurosciences, evolutionary biology, and philosophy. It can be broadly defined as two or more individuals coordinating their actions in space and time to achieve a particular joint outcome (e.g., Clark, 2006; Knoblich et al., 2011; see also Sebanz et al., 2006a). The common interest in joint action emanates from the growing body of research supporting its fundamental role for the emergence of the social mind, both phylogenetically and ontogenetically (Gallotti et al., 2017; Hasson et al., 2012; Milward and Carpenter, 2018; Tomasello, 2020a; Tomasello et al., 2005).

Each of the aforementioned research fields addresses fundamental aspects of joint action. For instance, neuroscientists seek to identify the brain areas and the neural processes involved in closely coordinated actions (e.g. in action anticipation, Denis et al., 2017; or brain-to-brain coupling, Gvirts and Perlmutter, 2020; Hamilton, 2021), and developmental psychologists study the ontogeny of joint action behavior (e.g. engaging in complementary roles, Meyer and Hunnius, 2020; understanding of joint commitments, Tomasello, 2020b). Further, evolutionary anthropologists and comparative psychologists focus on the evolution of joint action by studying nonhuman animals (henceforth animals, e.g. in social play, Heesen et al., 2021; coordinated decision-making, Duguid et al., 2020; or co-representation, Miss and Burkart, 2018). However, despite using the same umbrella term, namely joint action, there is considerable variation in the phenomena investigated in each of these fields.

Joint action is related to a number of similar concepts such as joint attention (e.g. Moll and Tomasello, 2007; Siposova and Carpenter, 2019), shared intentionality (e.g. Melis and Semmann, 2010; Tomasello et al., 2005), or joint commitment (e.g. Genty et al., 2020; Tomasello et al., 2005). A major caveat is that joint action and related concepts are notoriously difficult to conceptualize and operationalize. For instance, how could one decide whether two macaques engaging in a coalition against an opponent do so based on a shared goal? Current primate research may suggest to quantify mutual eye contact to establish joint commitment (e.g. Heesen et al., 2021), but the exchange of such communicative cues between interaction partners has also been described as a marker for joint attention (e.g. Siposova and Carpenter, 2019), action coordination (e.g. Bishop et al., 2019), and shared intentionality (e.g. Tomasello, 2018), making the distinction between these theoretical concepts difficult from an empirical perspective.

\textsuperscript{\textdagger} Corresponding author at: Department of Anthropology, University of Zurich, Switzerland.
E-mail addresses: fabia.miss@uzh.ch (F.M. Miss), judith.burkart@aim.uzh.ch (J.M. Burkart).

https://doi.org/10.1016/j.neubiorev.2022.104924
Received 30 March 2022; Received in revised form 11 October 2022; Accepted 20 October 2022
Available online 22 October 2022
0149-7634/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
The lack of conceptual clarity and a consistent terminology across disciplines hampers the possibility to build fertile interconnections and consequently, a proper integration of the findings across these fields has not been achieved yet. Instead, even closely related fields such as psychology and philosophy of mind have developed largely independent literatures (Milward and Carpenter, 2018). For instance, for many of the phenomena traditionally studied with neuroscientific approaches, the links between the processes studied there and ‘jointness’ as referred to by psychologists or philosophers remain unclear.

Given the emerging consensus that joint action and related concepts play a critical role for understanding human cognitive evolution, the question of how these phenomena are linked to each other – if at all – must have high priority. Such an alignment of several disciplines is a major endeavor, since working out even dyadic commonalities only, such as between comparative and developmental psychology (Heesen et al., 2017), or between psychology and philosophy (Milward and Carpenter, 2018), is far from straightforward.

A keen interest exists in explaining uniquely human hyper-cooperativeness and, thus, in clearly delineating whether animals do or do not engage in shared intentionality and related phenomena. Attempts to do so, however, tend to be hampered by the conceptual fuzziness of theoretical notions, turning them into moving targets for empirical scrutiny. A fruitful way out of this situation may be to directly ask what mechanisms animals and humans use when coordinating joint actions, and thus use an empirical bottom-up rather than a theoretical top-down approach to delineate the communalities and differences.

The goal of this review is to scrutinize a specific, empirical phenomenon, namely co-representation, which is arguably crucial for joint action, from the different theoretical and empirical perspectives. In doing so, we will ask how the phenomenon relates to key concepts in joint action research, and to what extent it should be construed as a low- or a high-level cognitive process. Co-representation provides a promising starting point for integrating results of joint action research from different fields because it can be studied using neurobiological, ontogenetic, and evolutionary approaches due to its amenability to adults, children, and animals.

Thus, we will start with a set of low- and high-level definitions used to describe joint action and related concepts, from simple, spontaneous motor coordination to cognitively highly demanding notions of shared intentionality, and show where co-representation is likely to fit in. We then describe standard behavioral paradigms to quantify co-representation and the most important factors influencing it. By reviewing the neurobiological underpinnings and what we know about the emergence of co-representation, ontogenetically and phylogenetically, and how it is linked to other (socio-) cognitive skills, including self-other (SO) integration and distinction, Theory of Mind (ToM) and inhibitory control, we will evaluate how cognitively demanding it is. We will conclude by highlighting how such a paradigm-driven bottom-up approach can help aligning joint action-related research across fields, and by flagging open questions and directions for future studies.

2. Joint action – theoretical perspectives

How can we describe different joint action phenomena, and how do they likely arise? Inter-individual coordination during joint action (e.g. Clark, 2006; Knoblich et al., 2011; Sebanz et al., 2006a) can take place in various contexts and rely on low-level, purely reflexive, automatic processes, or on high-level, cognitively demanding ones as those involved in processing others’ mental states. Thus, several socio-cognitive abilities such as SO integration and distinction or ToM may be more or less involved in regulating joint action, depending on how important it is to take the partner’s behavioral plans and mental states into account.

Low-level conceptualizations of joint action have also been referred to as emergent coordination (Fig. 1). It arises spontaneously, leading to similar behavior between individuals who have no plan to perform actions together (Knoblich et al., 2011), and does not involve common knowledge, or the understanding of another’s intentions. Knoblich et al. (2011) have identified four mutually non-exclusive sources that can lead to emergent coordination, namely (a) common affordances, where individuals perceive the same object at the same time, which specifies the action opportunities (e.g. a bench triggers sitting down on it, e.g. Gibson, 1977), (b) entrainment, i.e. the automatic temporal synchronization of the same or a different action (e.g. individuals spontaneously synchronize their hand-clapping rhythm, e.g. Schmidt et al., 2011), (c) perception-action matching, where observing an action activates the same action in the observer’s motor repertoire (e.g. observing someone dancing will activate corresponding action representations if one knows how to dance, e.g. Hommel et al., 2001), and (d) action simulation in common predictive models, i.e. matching observed actions onto the observer’s own motor repertoire enables an accurate prediction of action timing and outcomes (e.g. a dancer observing her partner initiating a particular jump may be able to predict the other’s position and timing of landing, e.g. Wolfert et al., 2003).

Emergent coordination can be observed in animals, as for instance in flocking in birds or schooling in fish. The mutual alignment in these groups is based on a few simple behavioral rules to move according to the position and orientation of the closest neighbors, resulting in group formation and cohesion (Ballarini et al., 2008; Cousin et al., 2002). Human automatic alignment may be observed during conversation, when speakers mutually adapt to each other’s speaking rate or imitate each other’s use of words, resulting in alignment of vocabulary and syntax (Hasson and Frith, 2016; Ruch et al., 2018). Other examples are people sitting in rocking chairs spontaneously synchronizing their rocking frequencies, or a couple falling into the same walking pace (Richardson et al., 2007; van Ulzen et al., 2008).

Goal-directed or planned coordination on the contrary is typically initiated internally, driven by the desired outcome of the joint action (i.e. jointly achieving a goal, Fig. 1). In the minimal case, goal-directed coordination requires an understanding of the joint action outcome, one’s own part in a joint action, and some awareness that the outcome can only be achieved with the support of another individual. Co-representation and joint attentional processes often support such goal-directed coordination (Knoblich et al., 2011). Co-representation is the mental representation of not only one’s own but also the partner’s actions (Ruisen and de Bruijn, 2016; Sebanz et al., 2006b, 2003; Vesper et al., 2010) and potentially also task rules specifying the conditions

---

**Fig. 1.** Mechanisms that can lead to emergent and goal-directed, planned joint action coordination, extended after Knoblich et al. (2011). Goal-directed, planned coordination can cognitively become increasingly complex if it additionally involves mechanisms involved at higher levels. The hierarchical model suggests that forms of coordination higher up in the pyramid build on mechanisms involved at lower levels. SO = self-other.
under which actions are to be performed (Sebanz et al., 2005a; Tsai et al., 2008; Vesper et al., 2013). Co-representation thus relies on processes responsible for achieving SO motor integration and distinction. SO distinction occurs when individuals (correctly) distinguish between themselves and the other (in the motor domain, e.g., imitative control, Santiesteban et al., 2012b; cognitive domain, e.g., false belief attribution, Decety and Lamm, 2007; affective domain, e.g., empathic feelings, Lamm et al., 2019). Self- and other-representations may merge when SO integration occurs without SO distinction (Colzato et al., 2012). Co-representation presumably facilitates joint action in that it enables individuals to better coordinate their actions and predict the joint outcome because both are anticipating and monitoring both sets of actions (Sebanz et al., 2006a; Sebanz and Knoblich, 2009; Sommerville and Decety, 2006). This leads to a higher precision and predictability of actions (Meyer et al., 2015). Moreover, integrating the co-actor’s actions (and perhaps also task requirements) into the own action planning can create a shared responsibility for the whole task when pursuing joint goals (Beyer et al., 2017; Butterfill, 2012; Sebanz et al., 2006a). Ideally, there is an optimal equilibrium between SO distinction and integration adjusted to the situation, to enable for instance smooth motor coordination between individuals (Bolt and Loehr, 2021; Köller et al., 2014; Vesper et al., 2017). Reaching such an optimal equilibrium requires control processes that permit individuals to mentally co-represent the partner’s actions while still being able to correctly distinguish between self- and other-generated actions (Knoblich et al., 2011). The control of one’s own actions and mutual adaptation to each other’s actions leads to a higher accuracy of actions in accordance with the co-actor (Meyer et al., 2015).

In addition to the partner’s actions, the individual may consider the other’s motives, perspectives, emotions, thoughts, knowledge, and beliefs. For instance, SO integration vs. distinction at this level may include empathy, to feel what the other is feeling but still being able to distinguish yourself from the other (Lamm et al., 2019, 2016). Likewise, joint action plans can be more refined and flexible if each other’s knowledge and intentions are considered. The partner’s motives, perspectives, emotions, thoughts, knowledge, and beliefs may also be negotiated to further optimize action plans or achieve long-term goals, which most likely requires language. Thus, the extent to which mutual representations of all these elements are taken into account during joint action can vary greatly during planned coordination (Knoblich et al., 2011; Vesper et al., 2010).

It is highly plausible that the types of coordination at the higher levels of the pyramid in Fig. 1 recruit mechanisms from the lower levels. For instance, spontaneous entrainment is likely to be involved as well when engaging in highly planned and negotiated shared activities. The list in Fig. 1 does not claim to be complete, and it is not evident for all mechanisms where they are best situated. Action co-representation for instance may be cognitively rather demanding. In fact, it has been reported to appear astonishingly late in ontogeny (e.g., measured with a joint Simon task, only in 4- to 5-year-old children, Milward et al., 2014; Saby et al., 2014, see below) and to depend on beliefs about the status of the partner as an agent. Specifically, in several studies, co-representation only emerged when coordinating actions with co-actors who were intentional or believed to be intentional (Ruys and Aarts, 2010; Sahai et al., 2019; Stenzel et al., 2012; Tsai et al., 2008; but see Wen and Hsieh, 2015). However, since it could likewise be a rather implicit mechanism directly emanating from perception-action matching, it is unclear whether co-representation per se is cognitively demanding, or whether co-representing specific contents (e.g., actions vs. perspectives and beliefs) is.

## 3. Joint action and co-representation: standard experimental paradigms and key results

As discussed in the previous section, co-representation is a crucial mechanism during joint action and arguably also for shared intentionality because it enables mutual action prediction and movement adjustment, and can facilitate a shared responsibility for the joint outcome. In this section, we will briefly highlight commonly studied joint action contexts and experimental designs that are likely to involve co-representation. Our major focus is on the joint Simon task, which has been extensively studied, and offers insights on the cognitive demands of co-representation as well as how the regulation between SO motor integration and distinction during dyadic action coordination may be achieved.

Studies on co-representation have mostly focused on dyadic tasks requiring the simultaneous or complementary coordination of identical movements. Examples are joint grasping, lifting, or pulling tasks (Constabile et al., 2021; Marshall-Pescini et al., 2017; Melis et al., 2006; Meyer et al., 2016; Newman-Norlund et al., 2008), synchronization tasks (Kirschner and Tomasello, 2009; Schmitz et al., 2017; Vesper et al., 2014, 2013; Visco-Comandini et al., 2015), or joint interference tasks such as the joint Simon task (Almada et al., 2011, 2008; Sebanz et al., 2005), see below. Other dyadic contexts requiring the simultaneous or complementary coordination of distinct movements likely also involve co-representation. These include for instance problem-solving tasks and social games (Cronin et al., 2005; Grafenbain et al., 2013, 2009; Kirschner and Tomasello, 2010; Warneken et al., 2006; Warneken and Tomasello, 2007), joint music performance (Liebmann-Jordanidis et al., 2021; Nivem et al., 2016), or social play (Heesen et al., 2021, 2017). Moreover, co-representation may be involved in tasks requiring the mental coordination of decision-making, such as in Stag Hunt, Egg Hunt, or puzzle box games (Duguid et al., 2020, 2014; McClung et al., 2017; Wyman et al., 2013).

Human studies suggest that the social context greatly influences the degree of SO integration, and thus likely co-representation, but also shared intentionality during joint actions. Studies on spontaneous synchrony for instance found that 2.5- to 4.5-year-old children synchronized their drumming tempo more accurately when drumming with a human partner than a drumming machine simulating a human hand, or a drum sound coming from a speaker (Kirschner and Tomasello, 2009), or that adult bodily synchrony was higher in affiliative than argumentative conversations (Paxton and Dale, 2013). In a joint Egg Hunt game, individuals engaged more in explicit goal sharing and joint action planning with in-group members than with out-group members, which led to increased cooperation success in the former (McClung et al., 2017). Reddish et al. (2013) looked at the interplay between shared intentionality and goal-directed behavioral synchrony. The authors found that explicit sharing of goals and intentions combined with synchronized body movements led to stronger SO integration (based on an adapted version of the Inclusion of the Other in the Self scale, Aron et al., 1992) as well as perceived unity, similarity, and trust (based on a questionnaire with 7-point Likert scales, Lakens and Stel, 2011; Wilternuth and Heath, 2009) than asynchrony or passive conditions. Interpersonal synchrony is further suggested to increase social bonding (e.g., Hove and Risen, 2009), commitment to cooperate (e.g., Cross et al., 2020; Kirschner and Tomasello, 2010), and may facilitate communication (e.g., Louwse et al., 2012). Interestingly, such a positive effect of synchrony on social outcomes can appear at low- and high- levels of joint action, since it has been described in emergent coordination (Hove and Risen, 2009; Richardson and Dale, 2005; Valdesolo and DeSteno, 2011), goal-directed coordination (Valdesolo et al., 2016; Wilternuth and Heath, 2009), and when sharing intentions (Reddish et al., 2013).

A joint interference task, the joint Simon task (Ruys and Aarts, 2010; Sebanz et al., 2003), has been repeatedly studied in humans, including investigations of underlying neurobiological processes, as well as in children and animals. It thus provides extensive data for studying joint action, in particular co-representation and the conflict resolution between SO integration and distinction when coordinating actions with a partner. Specifically, the joint Simon task exploits the situation in which co-representation hinders, instead of facilitates, joint performance. In
this task, pairs of participants complete a simple motor coordination task together. A typical task consists in correctly reacting to an external stimulus by choosing one of two response options, and each participant is responsible for one of them (see Box 1 for a description of the joint Simon task). This paradigm induces a strong tendency towards SO integration and co-representing the partner’s actions (quantified with the joint Simon effect, see Box 1), which in this case is detrimental to joint performance and thus cooperation success (i.e. the amount of correct response choices). In particular, co-representing the co-actor’s actions interferes with the individual’s own performance, and the individual fails to disambiguate between her own and the partner’s task affordances. Consequently, in the joint Simon task, suppressing co-representation is required to improve joint performance and, thus, maximize cooperation success.

Box 1
Assessing co-representation with the joint Simon task.

The Simon task was originally designed as an individual task (Simon and Rudell, 1967), with auditory or visual stimuli triggering a conflict in a motor response. For instance, subjects need to discriminate between two sounds by choosing a left-hand response option for sound “L” and a right-hand response option for sound “R” (Fig. 2). However, the sound itself is broadcast from either the left- or the right-hand side. Although the side of the broadcast is task irrelevant, subjects show more incorrect first orienting directions (i.e. body orientation or subtle move towards one side), incorrect manual choices, and longer response latencies when the sound is broadcast opposite to the correct response side (incompatible trial), than when the sides of the broadcast and the response match (compatible trial) (Simon effect). The Simon effect disappears when the subject only solves half of the task. Thus, when only one response option is available, but sounds “L” and “R” are still broadcast from either side, incompatible trials are no longer more difficult. Intriguingly however, the effect reappears when the second half of the task is solved by a partner (joint Simon effect). Thus, participants acting jointly on the task often demonstrate interference from the partner’s action/role in the task, even if this is irrelevant to their own task role (Sebanz et al., 2003). This suggests that the individual not only represents her own but also the partner’s actions (i.e. co-representation, e.g. Knoblich and Sebanz, 2006; Ruissen and de Bruijn, 2016; Sebanz et al., 2006b; Vesper et al., 2010). Therefore, co-representation can be measured experimentally at the behavioral level when it leads to interference on shared task performance.

Fig. 2. Experimental design of a joint Simon task. The task consists of four test conditions (a) full task, (b) half task, (c) joint task, and (d) joint-control task. In each trial, one of the auditory stimuli “L” and “R” is broadcast from one of the two lateral speakers, making it either a compatible trial (i.e. sound “L” and “R” are broadcast from the left-hand side and sound “R” and “L” from the right-hand side) or an incompatible trial (i.e. sound “L” broadcast from the left-hand side and “R” from the right-hand side). In monkeys, the manual response choice consists of pulling one of the two handles fixed on sliding drawers to retrieve a food reward out of the cup in case of a correct response choice. In the joint task, the partner subject can simultaneously retrieve the food reward from the cup in the middle. Miss and Burkart (2018) used this design to assess co-representation behaviorally in animals, providing evidence that marmoset monkeys indeed co-represent each other’s actions when performing the task together with a conspecific (see below for more details).

From the first application of the joint Simon task paradigm, a range of studies followed focusing on neurobiological processes, or the factors influencing co-representation, using different task designs (e.g. auditory or visual stimuli, social or non-social partner, see Table 1 and Fig. 3). Multiple human studies showed that the joint Simon effect could be modulated by certain social or socio-cognitive variables (e.g. McClung et al., 2013). Others found that the effect could also be elicited by the mere presence of an inanimate attention-getter, such as a metronome or a Japanese waving cat (e.g. Dolk et al., 2013). These studies therefore suggest either an inherently social or a purely perceptual, non-social account to explain the nature of the joint Simon effect. Thus, it is crucial to contrast the joint task condition with a joint-control condition in which a conspecific is present but cannot engage in the task (e.g. access to the response device is blocked, Miss and Burkart, 2018). In this control condition, all the low-level perceptual cues of a social partner are present, but the cooperation aspect is lacking to test the hypothesis of an inherently social origin of the effect.
Table 1
Joint Simon task studies suggesting either a referential coding account (green shading) or a co-representation account (blue shading) to explain the presence or absence of a joint Simon effect. The columns show: i. Study and type of stimuli used (visual or auditory), ii. Study subjects, iii. Covariation of the effect strength with the salience of attention getters or social (socio-cognitive) factors, and iv. Type of task (joint or independent) with instructions and salience of a joint goal (when mentioned). NT = not tested.

| Study and type of stimuli | Study subjects | Low-level spatial compatibility effect | Social effect Present? | Covaries with salience of social (socio-cognitive) factors Present? | Present despite no attention getter? | Type of task and joint goal salience |
|--------------------------|----------------|---------------------------------------|------------------------|-------------------------------------------------|--------------------------------------|-------------------------------------|
| Dolk et al., 2013         | Human adults   | YES: Japanese waving cat, clock, ticking metronome NO: silent metronome | Sense modality: no effect with silent metronome | NT: inanimate object | No co-actor |
| auditory see Fig. 2a      |                |                                        |                        |                                  |                                      | Joint task. Instruction to respond only to the sound assigned to the co-actor. Feedback (mean reaction time and percentage correct) at the end of a trial block (64 trials) |
| Dolk et al., 2011         | Human adults   | YES: co-author and proprioceptive event (rubber hand illusion (RHI)') YES: RHI and inactive co-author but see Miss and Burkart, 2018; Sebanz et al., 2003 | Sensory perception: effect smaller when co-authors received synchronous vs asynchronous stroking | Random co-actor |
| auditory                  |                |                                        |                        |                                  |                                      | Two independent half tasks. Both stimuli simultaneously presented on one screen |
| Guagnano et al., 2010     | Human adults   | Second actor performing an independent half task on the same screen within [YES] or outside [NO] arm-reach but see Welsh et al., 2013 | Spatial distance features: effect only if second actor within arm-reach | NT: second actor acting in parallel | Joint task. Instruction to refrain from responding to the non-assigned stimulus. Visual feedback on the screen in every trial |
| visual see Fig. 2b        |                |                                        |                        |                                  |                                      |                                      |
| Liepelt, 2014             | Human adults   | YES: co-author and response buttons' position either close to (on the monitor) or further away (on the knee) from the stimulus location | Spatial features: effect larger when co-author’s (rubber) hand is close to the stimulus location | Random co-actor | Joint task. Instruction to respond to only one color with the respective response move. Feedback exclusively for errors displayed only on the actor’s side where the error occurred |
| visual                    |                |                                        |                        |                                  |                                      |                                      |
| Dittrich et al., 2013, 2012 | Human adults | Response device as joystick [YES] or computer mouse [NO] YES: rubber hand only | Size: effect with joystick but not computer mouse | Random or same-gender co-actor | Joint task. Instruction to perform the task jointly with a confederate (headphones); Two independent half tasks (one visual and one auditory) in parallel on one device. Instruction to perform two separate tasks (headphones) |
| visual                    |                |                                        | Spatial features: effect only if spatial alignment of seating positions, response buttons and stimuli match | | | |
| Klempova and Liepelt, 2016 | Human adults | YES: framing of rare unexpected double (i.e. co-actors respond simultaneously) or single response events (i.e. only one actor responds at once) in a joint task or two independent half tasks | Unexpected rare events: effect larger after rare response events vs frequent ones | Random co-actor or second actor acting in parallel | Joint task. Instruction to respond to the assigned stimulus while a co-actor responds to the other one from a location either unknown or specified as opposite to the response button. Click sound signaling the co-actor’s executed response |
| visual or visual-auditory |                |                                        |                        |                                  |                                      |                                      |
| Sellaro et al., 2013      | Human adults   | YES: left computer mouse button if responding with the right one NO: co-author physically absent (position either known or unknown) see also Welsh et al., 2007; but see Ruys and Aarts, 2010; Tsai et al., 2008 | Everyday practice: effect only if responding with the right (less practice) vs left computer mouse button | Random co-actor | | (continued on next page)
### Table 1 (continued)

| Study and type of stimuli | Study subjects | Low-level spatial compatibility effect | Social effect | Type of task and joint goal salience |
|--------------------------|----------------|----------------------------------------|---------------|-------------------------------------|
| Sebanz et al., 2003      | Human (young) adults | NO: co-actor present but inactive | YES: random co-actor | Joint task (earplugs and covered hands) |
| visual                   |                |                                        |               |                                     |
| Hommel et al., 2009      | Human adults   | NO: intimidating, antagonistic co-actor | YES: friendly, supportive co-actor | NT Joint task. Instruction to perform the task jointly with a confederate. Feedback at the end of a trial block (60 trials) |
| visual                   |                |                                        |               |                                     |
| Ruys & Aarts, 2010       | Human adults   | NO: acting alone                      | YES: cooperative or (mildly) competitive co-actor in adjacent room | Joint task. Red light on the screen signaling every executed response. Instruction in one of three conditions: cooperate: rewarding both co-actors of the five best-performing dyads; compete: random selection of 10 winners (weak framing of competition) |
| auditory                 |                |                                        |               |                                     |
| Miss & Burkart, 2018     | Common marmosets (Callithrix jacchus) | NO: conspecific present but inactive | YES: tolerating conspecific | NT |
| auditory                 |                |                                        |               |                                     |
| Tai et al., 2008         | Human adults   | NO: robot described as functioning in a purely deterministic, machine-like way (passively controlled by a computer) | YES: believed intentional and acquainted co-actor in another room | NT Joint task. Instruction to cooperate with either a real person (acquainted prior to testing with possibility to communicate between trial blocks) or a computer in another room |
| visual                   |                |                                        |               |                                     |
| Stenzel et al., 2012     | Human adults   | NO: robot described as functioning in a purely deterministic, machine-like way (passively controlled by a computer) | YES: humanoid robot (active, intelligent) | NT Joint task. Visual feedback on the screen in every trial |
| visual                   |                |                                        |               |                                     |
| Ford & Aberdeen, 2015    | Human (young) adults | NO: robot described as functioning in a purely deterministic, machine-like way (passively controlled by a computer) | YES: unfamiliar or close-friend co-actor | NT Joint task. Instructions displayed for the dyad including the required action of both (participant A responding to one stimulus and participant B to the other) |
| visual                   |                |                                        |               |                                     |
| McClung et al., 2013     | Human (young) adult females | NO: ingroup co-actor (visual badge tagged to clothes categorizing group membership) | YES: ingroup (same-gender: female only) co-actor | NT Joint task. Instruction to respond to one color and withhold the response to the other. Feedback on accuracy after one trial block (8 practice and 64 test trials) |

(continued on next page)
| Study and type of stimuli | Study subjects | Low-level spatial compatibility effect | Social effect |
|--------------------------|---------------|--------------------------------------|---------------|
|                          |               | Is attention getter sufficient to result in effect? | Present? | Covaries with salience of social (socio-cognitive) factors? | Present despite no attention getter? | Type of task and joint goal salience |
| Müller et al., 2011b visual | Human (young) adults | NO: outgroup co-actor (displayed on the monitor as animated hand with either light or dark skin pressing the response button) | YES: ingroup co-actor displayed as animated hand | Group membership (categorization with light- or dark-colored hand); effect only with ingroup vs outgroup co-actor; however, effect with outgroup co-actor after explicit instruction of perspective taking see also McClung et al., 2013 | NT | Joint task. Instruction to respond to only one color with the respective button |
| Kuhbandner et al., 2010 visual | Human adults | NO: co-actors in a negative mood state | YES: random co-actor in the same neutral or positive mood | Affective state (mood): effect absent when co-actors in a negative, present in a neutral, and stronger in a positive mood state<sup>a</sup> | NT | Joint task. Instruction to respond to either stimulus A or B |
| Ruissen and de Brujin, 2015 visual | Human adult males | NO: co-actor when the task is preceded by a competitive joint flanker task (JFT) | YES: explicit cooperative co-actor after a competitive JFT | Carry-over effect of competition: no effect when preceded by an unrelated JFT under competitive instructions; see also Ruissen and de Brujin, 2016; effect weaker after competitive Tetris game play | NT | Joint flanker task (JFT) under instructions to either cooperate (rewarding the best-performing dyad) or compete (rewarding the best-performing participant), ▪ Follow-up joint Simon task with explicit instructions to cooperate (rewarding the best-performing dyad) |
| Iani et al., 2014 visual | Human adults | NO: random (same-gender: male only) co-actor | YES: no auditory perception of the co-actor’s actions | ▪ 2 – 5-year olds: Non-spatial joint task. Cardboard screen separating the children. Instructions to each child separately given by a puppet. Positive feedback and reward (sticker) at the end of the experiment |
| Milward et al., 2014 visual | Human children (2 – 5-year old) | YES: random 4 – 5-year-old child or experimenter as co-actor | Age and linked ToM: effect in 4 – 5-year-olds (non-spatial joint task); one of two stimuli appears in the center and performance when participant X responds to stimulus A and participant Y to stimulus B is compared to performance when both respond to the same stimulus), but no effect in 2 – 3-year-olds see also Saby et al., 2014; effect in 5-year-olds with the experimenter as co-actor; but see Miss, 2022: effect in 2 – 4-year-olds tested with primate task design | NT | ▪ 2 – 5-year olds: Non-spatial joint task. Experimenter as co-actor. Reminder of instructions and positive feedback after one trial block (12 trials) |

<sup>a</sup> RHI is induced by the synchronous stroking of a stroking device while fixating the eyes onto the co-actor’s hand which can trigger the illusion of the co-actor’s hand becoming a part of the own body.

<sup>b</sup> A revised version of “The mind in the eyes test” (Baron-Cohen et al., 2001) was used that required the participants to select one of four terms (e.g. serious, ashamed, alarmed, bewildered) that best described the emotional state of different pairs of eyes (for a total of 36 facial expressions).

<sup>c</sup> EQ (Lawrence et al., 2004) is based on a self-report measure using 60 rating-scale questions that result in an overall score on cognitive perspective taking (i.e. ToM), emotional contagion, and social skills. IRI (Davis, 1980) is based on a self-report measure using 28 rating-scale questions that result in four separate scores (sympathy, perspective taking (i.e. ToM), fantasy, and personal distress).

<sup>d</sup> The degree of interpersonal closeness between co-actors was assessed with a modified version of the ‘Inclusion of the other in the self (IOS) scale’ (Aron et al., 1992) resulting in four scores: neutral, slightly close, moderately close, and extremely close.

<sup>e</sup> A mood state was induced prior to the joint Simon task with a film clip presented according to a published library of films serving the induction of mood (neutral: “Sticks”, positive: “Harry and Sally”, negative: “The Champ”). The success of mood induction was assessed with an affect grid (a nine-by-nine matrix varying along the dimensions of valence (extremely negative to extremely positive) and arousal (low to high).
Studies that have been investigating co-representation based on interference in joint action tasks across ontogeny and species provide an intriguing puzzle. Co-representation has been found in adults and in 4- to 5-year-old children, but not in younger ones (Milward et al., 2014; Saby et al., 2014), which may suggest the involvement of high-level cognitive processes linked to ToM and inhibitory control abilities (see paragraph 4.2 on ontogeny). However, co-representation is also present in marmoset monkeys (Miss and Burkart, 2018), capuchin monkeys, and Tonkean macaques (Miss et al., 2022a) (see paragraph 4.3 on phylogeny). None of these three species has ToM abilities comparable to 4- to 5-year-old children, and also their inhibitory control abilities are clearly inferior (MacLean et al., 2014). Solving this puzzle requires us to scrutinize the potential mechanisms and modulating factors of the joint Simon effects in adult humans, as well as the emergence of co-representation at the neurobiological, ontogenetic, and phylogenetic level.

3.1. Potential mechanisms and modulating factors of the joint Simon effect

Before analyzing the underlying neurobiological processes and the emergence of co-representation during development and evolution in more detail, we take a step back in this section, and review two prominent accounts, co-representation and referential response coding, to explain the nature of the joint Simon effect. In particular, we evaluate task environments in which co-representation is most likely to emerge in the joint Simon task, namely, in social task set-ups and when actor interdependence is made salient.

The joint Simon effect has been explained to origin from genuinely social co-representation, but alternatively also from non-social referential response coding (for a review see Dolk et al., 2014), which gave rise to some controversy. In this section, we will argue that much of this controversy can be resolved when taking the details of the applied task designs into account. Table 1 gives an overview over joint Simon task (-like) experimental studies which explained their results either in favor of the referential coding or the co-representation account. It also lists the factors which appear to be highly relevant for the outcome, namely the salience of social factors, non-social attention getters, and a prominent joint goal.

3.1.1. Potential mechanisms

The referential response coding approach (Dolk et al., 2013) describes the joint Simon effect as based on the spatial components of the task itself, and on purely perceptual mechanisms. Therefore, the (social or non-social) co-actor is used as a spatial reference point to code one’s own actions spatially as either left or right. Any attention-grabbing event (e.g. a Japanese waving cat, see Fig. 3a) would require participants to discriminate between events that are self-controlled and events that are not. The discrimination problem between internally activated and externally produced events then creates a joint spatial compatibility

Fig. 3. A selection of four prominent task designs used to test subjects with the joint Simon task. The setup can contain (a) auditory stimuli and a non-social co-actor (e.g. a Japanese waving cat, giving rise to the referential response coding approach and the spatial compatibility effect), (b) visual stimuli and co-actors performing two independent half tasks in parallel (spatial compatibility effect), (c) auditory stimuli and a joint reward structure (social joint effect), and (d) visual stimuli and either a human (social joint effect) or a non-biological agent (e.g. a computer; no social joint effect) as believed co-actor in a different room.
effect (joint Simon effect). Consistent with this approach are findings that show how a mere attention getter or externally produced event, instead of a social partner, can trigger the effect, independently of the task of the co-actor (Dittrich et al., 2013, 2012; Dolk et al., 2011; Guagnano et al., 2010; Klempova and Liepelt, 2016; Lien et al., 2016, see Fig. 3a–b). However, in nonhuman primates (henceforth primates, and see Sebanz et al., 2007, 2003 for the same finding in humans), a conspecific being present, next to the individual performing the task, without engaging itself in the joint activity does not elicit co-representation as observed when both partners solve the task together (Miss and Burkart, 2018, see Box 1). The co-representation account explains the joint Simon effect as originating from the mental representation of one’s own and the partner’s actions and perhaps also task rules governing these actions in a functionally equivalent way. Thus, each actor integrates the co-actor’s alternative action into her own action planning and behaves as if she was not only responsible for her own but also her partner’s actions. The co-representation account thus proposes that the mechanism originates from the social context and the interaction component when performing the task together with a partner (Kiernan et al., 2012; Knoblich et al., 2011; Ruisen and de Bruijn, 2016; Sebanz et al., 2006; Vesper et al., 2010, see Fig. 3c–d).

### 3.1.2. Modulating effects: social factors and goal salience

Social factors can modulate the strength of co-representation, as measured in joint Simon tasks. For example, individuals tend to show a weaker or no effect in a competitive than cooperative context (Iani et al., 2014: Ruisen and de Bruijn, 2016) or with an outgroup than ingroup co-actor in case of salient group categorizations (McClung et al., 2013; Müller et al., 2011b; see Iani et al., 2011 for no modulation with a minimal group categorization). Other examples show weaker or no effects after the induction of a negative than positive affect between the co-actors (Kuhlbander et al., 2010), or with an antagonistic instead of supportive co-actor (Hommel et al., 2009). Further, the strength of co-representation positively correlates with cognitive empathy (i.e. perspective-taking) among friends (Ford and Aberdein, 2015) or with the degree of interpersonal closeness between co-actors (Safaie et al., 2020). Moreover, the effect increases after nasal oxytocin application (Ruisen and de Bruijn, 2015) (oxytocin is a neuropeptide known to affect social preference and partner-directed social behavior in humans, Heinrichs et al., 2009; rodents, Insel and Young, 2001; and primates, Smith et al., 2010). Further, co-representation also emerges with an invisible co-actor that is intentional, or at least believed to be intentional. Here, it seems necessary and sufficient that the partner believes that the responses are from an intentionally acting partner rather than automatically generated by an algorithm (Ruys and Aarts, 2010; Sahai et al., 2019; Tsai et al., 2008, see Fig. 3d). In sum, there is considerable evidence that social factors may modulate co-representation. This strengthens the social co-representation account of action coordination rather than the referential response coding approach. However, it is likewise possible that the modulation by social factors is driven by an attentional bias in that subjects might pay more attention to partners with whom they have a closer relationship (i.e. familiarity bias) and therefore, the effects of referential response coding would be stronger with more strongly bonded or familiar partners.

The referential response coding approach and the social co-representation approach can be reconciled in two ways. One is to assume that non-social, referential mechanisms are always at work, but sometimes supplemented by social co-representation. This alternative is unlikely however, since under some conditions, joint effects have not been found despite the presence of strong attention getters (Miss and Burkart, 2018; Sebanz et al., 2007, 2003; and all the studies with modulating social variables). Further, under some conditions joint effects have been found despite the absence of attention getters (Ruisen and de Bruijn, 2015; Ruys and Aarts, 2010; Sebanz et al., 2003; Tsai et al., 2008). Likewise, the referential coding account cannot explain findings showing an effect only with (believed) intentional compared to non-intentional co-actors (Sahai et al., 2019; Stenzel et al., 2012, 2014; Tsai et al., 2008, see Table 1).

Another possibility is that the two mechanisms are activated depending on some external factors, namely task settings and instructions, which indeed can readily explain some contradictory findings in the literature. The most important of these factors appears to be the extent to which the task design makes a joint goal salient and provides a social context that creates interdependence, making both individuals’ contributions of comparable importance to reach the joint outcome and, thereby, favoring joint agency attribution (Pacherie, 2012). For instance, whether the goal is perceived to be shared likely differs when actors share the same task with a co-actor (e.g. Ford and Aberdein, 2015; Hommel et al., 2009; Iani et al., 2014; Tsai et al., 2008), and when they perform separate tasks simultaneously next to each other (e.g. Guagnano et al., 2016; Klempova and Liepelt, 2016; Milward et al., 2014). Whether participants perceive that they are pursuing a joint goal can be influenced explicitly. For instance, subjects can be verbally instructed that they are expected to solve a cooperation task together (e.g. Hommel et al., 2009; Tsai et al., 2008) or that an intentional (i.e. goal-directed) cooperation partner is in a different room (e.g. Ruys and Aarts, 2010; Sahai et al., 2019; Tsai et al., 2008). These examples clearly differ from instructions that focus exclusively on individual performance to respond to only one stimulus (e.g. in Dittrich et al., 2012; Dolk et al., 2011; Müller et al., 2011b). A joint goal can also be made salient non-verbally, by rewarding both partners after a correct answer regardless of which actor provided it (e.g. Miss and Burkart, 2018, see Fig. 3c). In the joint condition of this study with common marmosets, both individuals could act on the task, and both received a reward in case of a correct choice. In a joint-control condition, on the contrary, the partner subject was present but could neither contribute to the task nor did it receive a reward if the actor chose correctly. The mere physical presence of a social partner (i.e. a very strong attention getter) was not sufficient to elicit the effect in marmosets unless the partner was jointly engaged in the same task (as in humans, see Sebanz et al., 2007, 2003; but see Dolk et al., 2011). Moreover, the salience of a joint goal can also be influenced implicitly. For example, the same visual or auditory feedback signal for a correct response can be provided to both the actor and the partner in each trial (e.g. Liepelt, 2014; Stenzel et al., 2012). This task design is different from a summarized feedback at the end of a trial block or the experiment (Dolk et al., 2011; Hommel et al., 2009; McClung et al., 2013; Milward et al., 2014). Further, a joint goal and interdependence might become less salient if error feedback is visible only to the actor but not to the partner (Dittrich et al., 2012), or when co-actors cannot see each other (e.g. co-actors separated by an opaque cardboard screen, Milward et al., 2014, see Table 1). To get rid of the confounding factor of task instructions and find ways to comparatively study co-representation, such as in children with less sophisticated language skills than adults or in animals, a joint reward structure may be a particularly fruitful task feature when testing subjects with the joint Simon task.

In sum, as shown in Table 1, referential response coding can be the explaining mechanism for the joint Simon effect in non-social task settings. However, it cannot explain why co-representation emerges in a joint but not a joint-control condition (see Box 1). Moreover, it cannot easily be reconciled with the abundant evidence in social task settings suggesting that co-representation is modulated by a variety of social and socio-cognitive factors, such as social distance, relationship quality, cognitive empathy, or even whether the co-actor actually performed the action intentionally (see above, e.g. Welsh et al., 2013; but see Hommel, 2019). These findings suggest that, as soon as the co-actor is a social partner, the effect involves more than merely bottom-up driven reactions to perceptual inputs from the environment (e.g. Dolk et al., 2013). Overall, it appears that the most relevant factor for a true social effect to emerge, i.e. if the partner’s actions get co-represented or not, is whether the task is perceived to be shared or not (e.g. Sebanz et al., 2006a). Therefore, co-representation appears most involved in joint task set-ups which
increase the salience of a shared goal. In human adults, to create actor interdependence and the perception of a joint goal, it is likely sufficient to provide a simple verbal instruction, such as to solve a cooperation task together with a partner, or a feedback signal indicating success/error per trial to both partners. In comparative studies, including (young) children and animals, this may similarly be achieved with a joint reward structure (i.e. rewarding both individuals in case of a correct response choice).

4. The emergence of co-representation

To further pinpoint the suggested social nature of co-representation, and to what extent it should be construed as a lower- or higher-level cognitive process (Fig. 1), we will next review its neurobiological underpinnings and how it emerges during ontogeny and phylogeny.

4.1. Neurobiological findings

Social neuroscientists seek to identify the neural mechanisms that allow individuals to closely coordinate their behavior. This is typically studied in human adult subjects, most commonly with the help of electrophysiological tools to analyze electroencephalograms (EEGs), event-related potentials (ERPs), lateralized readiness potentials (LRPs), event-related functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS). These approaches study action simulation, anticipation, monitoring, brain-to-brain coupling, etc., in joint action paradigms involving similar motor behaviors, such as during joint bar lifting (Newman-Norlund et al., 2008), joint rhythmic behavior (Novembre et al., 2016), or the joint Simon task (Ruisen and de Bruijn, 2015; Sebanz et al., 2007).

4.1.1. Simulation-, ideomotor-, and mirror neuron theory and their potential shortcomings in explaining social interactions

Several studies suggest that individuals internally simulate actions performed by an interaction partner (simulation theory, Gallese and Goldman, 1998) and integrate this simulation with representations of own action goals during joint action (Bekkering et al., 2009; Sebanz et al., 2007, see also common predictive models in Fig. 1). Importantly, the neural simulation of actions performed by an interaction partner seems to rely on the familiarity (or internalized knowledge) and sensorimotor experience with a given task (Hadley et al., 2015; Novembre et al., 2016). A part of this process likely relies on the brain’s capacity to code action production (related to the self) and perception (mostly related to others) in a comparable way (i.e. perception-action matching, see Fig. 1), such that actions and their outcomes can be maintained in a common representation (Rizzolatti and Sinigaglia, 2010).

Ideomotor approaches (or common coding, Hommel et al., 2001; Prinz, 1997) investigate how observing others’ actions can induce a tendency to engage in these actions ourselves. According to ideomotor theory, perceiving events produced by another’s actions should activate the same representational structures underlying own planning and control of these actions (Jordan and Knoblich, 2004). Therefore, in the context of joint action, it seems plausible to represent perceived motor behavior of another person within one’s own motor repertoire and incorporate it in the planning of subsequent actions (Tsai and Brass, 2007, see SO integration in Fig. 1). Mirror neurons, originally discovered in the macaque brain, fire both when the macaque executes an action and when it observes the same act performed by others (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996). Mirror neurons have further been documented in marmosets (Suzuki et al., 2015) and songbirds (Keller and Hahnloser, 2009; Prather et al., 2008). With the difficulty of single-cell recordings in humans there is very little direct evidence of mirror neurons in humans, however, neurons with mirror neuron properties also seem to exist in the human brain, providing indirect evidence of a mirror neuron system or circuit in humans (see Rizzolatti and Craighero, 2004). Thus, the generation of a cortical representation of an observed action in the motor systems of the observer represents a plausible neural mechanism for integrating one’s own and observed actions into a common representation (Gallese et al., 1996; Rizzolatti et al., 2001).

Even though mirror neurons may play a role in action understanding (Cook et al., 2014; Rizzolatti et al., 2001; Umiltà et al., 2001), a sole bottom-up “mirroring” of action goals when coordinating actions with others seems unlikely since there is no one-to-one mapping of perceived movements and the goals associated with them (Lamm and Majdandžić, 2015). Therefore, independent work on human social cognition has brought forward discussions on whether the mirror neuron activation itself may function as egocentric system which does not include the social interactional component inherent in joint action, raising the question whether it is sufficient to account for cooperative behavior in the context of how individuals manage to act together (Knoblich and Jordan, 2003; Meltzoff and Decety, 2003; Pacherie and Döns, 2006; Sebanz and Knoblich, 2009; Tsai et al., 2008; but see Blakemore and Decety, 2001; Blakemore and Frist, 2005). Similarly, ideomotor theory as described in the context of processes for motor programming within - rather than between - individuals may not be sufficient to explain flexible behavioral changes in a dynamic social context that requires individuals to perform different actions side-by-side, take turns, and coordinate their behavior to reach a joint goal (Sebanz et al., 2006a, 2003). The understanding of the functions and limitations of mirror neurons is part of ongoing research, and there is no overall agreement on whether mirror neuron activation reflects an action (i.e. other brain areas are responsible for making inferences about goals and intentions and mirror neuron responses generate a predictive simulation to facilitate ongoing action perception, Wilson and Knoblich, 2005), or indeed indicates understanding of an action or the action-related mental states of others (e.g. Hickok, 2009; Molenberghs et al., 2009; Saxe, 2005). Thus, on the one hand, mechanisms such as perception-action matching and SO motor integration originating from simulation-, ideomotor-, and mirror neuron theory appear sufficient for explaining forms of emergent coordination. On the other hand, in case of goal-directed, planned coordination, they are likely to be involved but interacting with, and supplemented by, additional mechanisms (Fig. 1).

4.1.2. The relevance of inhibition processes in action control, anticipation, monitoring, and attribution of agency

Electrophysiological and fMRI studies on joint Simon (-like) tasks show control processes are involved to inhibit performing an action when it is the other’s turn (Knoblich et al., 2011), which is overall consistent with the need to suppress SO integration in order to achieve SO distinction and perform well. The electrophysiological component P3 (a positive event-related potential occurring 300 – 500 ms post-stimulus) is suggested to be an index of action control and response inhibition in no-go trials (Falkenstein et al., 2002). In ERP studies on the joint Simon task, a larger P3 was observed in trials requiring the coactor’s response in the joint task condition (i.e. no-go trial in the joint task) than in no-go trials of the half task condition. Larger P3 in no-go trials in the joint Simon task might therefore indicate increased response inhibition to prevent one from responding when it is the partner’s turn because one’s own and the partner’s actions are formed in a common representation (Ruisen and de Bruijn, 2015; Sebanz et al., 2006b; Tsai et al., 2006, see co-representation and balancing of SO integration vs. distinction in Fig. 1). Further, it might reflect increased action anticipation for a better control of action monitoring when acting together on the task (Tsai et al., 2006). Tsai et al. (2006) additionally analyzed the lateralized readiness potential (LRP, time-locked to the stimulus onset), an ERP-component indicating the preparation of motor responses to a stimulus, such as when the task requests either left- or right- hand responses (i.e. response selection and activation of action planning, Coles, 1989). They found the LRP from compatible no-go trials and incompatible go trials to be larger in the joint task condition than in
the half task condition. The authors suggested a priming of cortical responses such that own response activation occurs with the anticipation of the partner’s actions. This then leads to interference at the response selection stage no matter who’s turn it is (Tsai et al., 2006, see also common predictive models in Fig. 1). Similar results were obtained in a turn-taking task using only centrally presented stimuli eliciting no interference effect. LRPs were found to be present in no-go trials in the joint condition when it was the partner’s turn to respond, but they were absent in no-go trials of the half task condition (Hollandér et al., 2011).

Interestingly, in a joint Simon task with explicit instructions to cooperate, increased action control and response inhibition in partner trials was observed only with an unseen and intentional human agent as co-actor, but not with a computer situated in a different room (Tsai et al., 2008, see Fig. 3d). The generation of predictions of the other’s actions, based on the representation of the other’s task and action anticipation, was also observed in a similar associative joint stimulus-response task, in which visual cues specified future actions of either of two co-actors in adjacent rooms (Ramsani and Mirolli, 2004). Neural activation related to the anticipation of the co-actor’s responses was observed in areas outside the classic mirror system (ventral premotor cortex, the paracingulate cortex and superior temporal sulcus), which have been found to be activated as well during mental state attributions (Fridh and Fridh, 1999; Gallagher and Fridh, 2003, see also higher-level representations of minds in Fig. 1). Being involved with other people’s actions has been considered to induce uncertainty about agency (i.e. who is performing the perceived action, Beyer et al., 2017; Colzato et al., 2012; Stenzel et al., 2012). During dyadic music performance, neural activity in the right centroparietal brain area (Novembre et al., 2016, see below for details) showed an overlap with brain areas involved in sense of agency (anterior medial prefrontal cortex (aMPC) and the right temporoparietal junction (rTPJ)). The latter area is suggested to be responsible for relating own body movements to other’s movements (Brass et al., 2005), and to be involved in visual perspective taking (Ruby and Decety, 2001) and attribution of agency (Blakemore and Frith, 2003; Decety and Lamm, 2007, see balancing of SO integration vs. distinction in Fig. 1).

In an fMRI study, Sebanz et al. (2007) applied a joint Simon task condition and a half task condition in which the participant responded to the assigned stimulus, while ignoring the other stimulus, side-by-side with an inactive confederate who only rested her finger on her response key. The latter task condition thus strongly resembled the joint-control task condition of the marmoset study (Miss and Burkart, 2018, Box 1) in which a conspecific was present but not engaged in the task. The study showed increased neural activation in the joint but not the half task condition when participants acted upon stimuli that required their own response. In particular, the activation occurred in the ventral MPC, an area suggested to be engaged in different tasks, and linked to self-referential processing (i.e. SO distinction, Mitchell et al., 2006), joint attention (Williams et al., 2005), visual perspective taking (Vogelley et al., 2004) and thinking about others’ beliefs and intentions (Amodio and Frith, 2006 for a review). The authors hypothesized that representing the partner as a potential actor increased the relevance of the stimuli referring to oneself. Joint performance with a co-actor was further associated with increased orbitofrontal activation, indicating a closer monitoring of the performances to make sure that when responding, it really was their turn (see joint attentional processes in Fig. 1). In trials requiring the partner’s response, the results further indicated an increased activation in the parietal lobe and the supplementary motor area suggesting it indices inhibition of motor responses (Durston et al., 2002), and might therefore reflect similar results as the ones obtained in ERP studies (Sebanz et al., 2006b; Tsai et al., 2006). In further fMRI research, Wen and Hsieh (2015) applied a belief paradigm (as in Tsai et al., 2008), in which the participant performed the joint Simon task with a believed human co-actor or a computer located outside the scanning room. They found the activation in compatible trials in the MPC to be higher in the human co-actor condition than in the computer condition, and the authors therefore hypothesized that this increased activity reflects SO distinction when believing to interact with a biological co-actor (in line with arguments from Sebanz et al., 2007).

Thus, with the likely involvement of low-level mechanisms of SO motor integration, perception-action matching, and action simulation in co-representing the partner’s actions during action coordination, co-representation may emerge rather automatically. The recruitment of inhibition processes linked to increased action control, and activation of structures involved in agency attribution and action monitoring in the joint Simon task, is further in line with the given task characteristic, namely that co-representation is detrimental to task performance. In this particular situation, it is thus best suppressed to resolve conflict between SO integration and distinction and correctly differentiate between self- and other-generated actions.

4.1.3. Self-other integration and distinction, and its regulation during brain-to-brain coupling

SO integration and distinction of motor actions may also be linked to SO integration and distinction of mental states (the highest level in Fig. 1). Experimental and correlational evidence, based on other paradigms than the joint Simon task, suggests that motor SO integration facilitates a broader common representational basis for action coordination involving long-term joint planning and shared mental states (Hasson and Frith, 2016; Kampis and Southgate, 2020; Keller et al., 2014). In particular, SO integration is suggested as crucial component in simple motor coordination (e.g. Brass and Heyes, 2005), but also more complex forms of joint action relying on a joint focus of attention (Bökler et al., 2012), memory performance (Ekenazi et al., 2013), or visual perspective-taking (in adults, Samson et al., 2010; in 6-year-olds, Surtees and Apperly, 2012). Further, it seems to be involved in egocentrically biased affective judgments (emotional egocentricity bias, i.e. the tendency to egocentrically attribute the own emotions to the other, Silani et al., 2013), consistent with the suggestion that an understanding of others’ thoughts and feelings partly relies on projections of own experiences (Lamm et al., 2019; Steinbeis, 2016).

That motor SO distinction indeed shows corresponding involvement in these skills has been shown by Santieteban et al. (2012b). In a task that required attributing conflicting perspectives (i.e. the Director’s task), the training to inhibit imitation, but not of motor inhibitory control in general, increased the participants’ subsequent performance in the task. This suggests processes of SO distinction in the motor domain (and not only executive functions per se such as inhibitory control ability developing rather late during ontogeny) to be relevant for visual perspective-taking. Moreover, neuroimaging studies suggest a partial overlap in neural activity in the rTPJ when distinguishing between self- and other-generated actions (e.g. Brass et al., 2009), comparing internal expectations with external events (e.g. Decety and Lamm, 2007), processing gaze (e.g. Hamilton, 2016 for a review), and attributing mental states (e.g. Saxe and Kanwisher, 2003; Spengler et al., 2009). This brain area has therefore repeatedly been associated with SO distinction in motor (e.g. imitative control) and cognitive (e.g. mental state attribution) domains (e.g. Decety and Lamm, 2007; Santieteban et al., 2012a).

SO distinction is also required to overcome emotional egocentricity and to empathize with others, thus separating empathy (i.e. feeling as, and understanding, the other with SO distinction) from emotional contagion (emotion-matching without SO distinction) and personal distress (a self-oriented, aversive emotional response, Adrienne et al., 2020; Bukoski et al., 2020; Lamm et al., 2019).

An intriguing recent approach gaining growing interest is offered by brain-to-brain coupling studies that focus on investigating potential interactions between the neural mechanisms of joint action partners’ brains when engaged in real-time joint action (e.g. Hasson et al., 2012). For example, in a musical dyadic joint action paradigm, Novembre et al. (2016) found the mechanisms underlying SO integration and distinction and co-representation to be linked between the partners through neural rhythms originating from multiple brain regions, including
somatosensory, motor, and parietal areas. Such neural oscillations have been suggested to be involved in interbrain processes, but their distinct roles still need further evaluation (Djolovski et al., 2021; Sedley et al., 2016). Alpha-oscillations (8 – 12 Hz) are thought to sustain the construction of predictions, beta-oscillations (13 – 30 Hz) the accuracy of predictions, and gamma-oscillations (31 – 48 Hz) are thought to be involved in predicting errors and refining action through sensory input (Djolovski et al., 2021). In November et al.’s study (2016), pairs of pianists performed short complementary musical duets while action familiarity and interpersonal synchronization accuracy were manipulated. Only when the pianists were familiar with each other’s parts did higher temporal synchronization (induced by tempo congruence between the two players) lead to alpha-suppression, which then favored SO integration and greater mutual adaptation. Lower temporal synchronization (tempo incongruence between the two players) on the contrary led to alpha-enhancement which then favored SO distinction. Alpha oscillations might therefore be involved in regulating the balance between SO integration and distinction depending on the compatibility of internal (knowledge) and external (environmental) information during motor coordination tasks (Novembre et al., 2015, see balancing of SO integration vs. distinction in Fig. 1). Consistent with this finding, other dual-EEG studies showed that socially interactive tasks requiring temporal motor coordination (e.g. joint rhythmic behavior, joint speech, imitation of hand movements) are associated with pools of neurons oscillating coherently across co-actors’ brains (Keller et al., 2014 for a review). In particular, the accuracy of synchronization seems to result from error correction processes that modulate the coupling strength between internal timekeepers and the external pacing signal (e.g. generated by a computer or an individual), and the degree of mutual adaptation between interacting partners (Hasson and Frith, 2016; Repp and Su, 2013; van der Steen et al., 2015). In an fNIRS hyperscanning study by Piazza et al. (2020), brain-to-brain coupling between adult caregivers and one-year-old infants was higher during direct interactions (playing, singing, reading) than when each engaged in the same interactions with another person. Moreover, moment-to-moment social dynamics, such as mutual eye contact, infant smiling, or joint attention to an external object significantly contributed to neural alignment in these caregiver-infant pairs. Further, Djolovski et al. (2021) targeted the interplay of neural and behavioral synchrony in a goal-directed, complementary motor task (“Etch A Sketch”) in adult dyads with varying relationship status from being unfamiliar to long-term couples. Using hyperscanning EEG, they found interbrain connectivity in the joint motor task in beta and gamma oscillations localized to sensorimotor areas. Interestingly, they found the highest interbrain synchrony and behavioral synchrony (e.g. reciprocity of interaction, fluency and rhythmicity of the interaction, mutual adaptation of the two partners), and therefore most efficient performance (i.e. time needed to reach the shared goal), in long-term couples compared to friends and strangers. Thus, observed influences of social factors, such as increased familiarity between joint action partners, on behavioral synchrony may be reflected in synchronous interbrain activity, which can further be linked to unique, dyad-specific interaction dynamics like mutual eye contact or joint attention. To what extent synchronized brain oscillations across interaction partners are linked to SO integration and distinction at higher cognitive levels (i.e. of perspectives, emotions, beliefs) will be an important endeavor in future research.

In sum, individual measures suggest that the neural mechanisms allowing individuals to closely coordinate their behavior involve processes that are responsible for coding action production and perception in a comparable way (e.g. Rizzolatti and Sinigaglia, 2010, see Fig. 1). Relying on the familiarity and sensorimotor experience with the task (e.g. Hadley et al., 2015), individuals in the joint Simon task seem to internally simulate actions performed by the co-actor and integrate this simulation with representations of own action goals and planned subsequent actions (Sebanz et al., 2007, see Fig. 1). Mirror neurons and action simulation, however, may not explain all the neural mechanisms important for joint action and achieving a shared goal (Hickok, 2009; Sebanz et al., 2006a). In particular, motor inhibitory control processes also seem to play an important role in joint Simon task-like paradigms, preventing one from acting when it is the partner’s turn to respond, and increasing action anticipation and monitoring (e.g. Ruisen and de Brujin, 2015; Tsai et al., 2006, see Fig. 1). Studies on interbrain processes, such as brain-to-brain coupling during real-time joint actions (e.g. joint rhythmic behavior) show that co-representation is linked to neural oscillations in co-actors’ brains, regulating the balance between SO integration and distinction, which allows to flexibly control and adjust actions to one another (e.g. Novembre et al., 2016, see Fig. 1). They are thus a valuable tool for future studies to address SO integration also at higher conceptual levels, in particular because it seems likely that SO integration and distinction of motor actions form the basis for SO integration and distinction at higher-level representations of minds that involve for instance perspectives, emotions, or beliefs.

Interindividual neural variation also appears modulated by social context. This further supports the view that the joint Simon effect is a genuinely social effect originating from the interactional component of performing the task jointly and from co-representing the partner’s task and actions (e.g. Sebanz et al., 2006a), rather than an effect emerging from purely bottom-up driven reactions to any (non-social or social) environmental cue (e.g. Dolk et al., 2013, see paragraph 3.1.1 and Fig. 1). In conclusion, it appears that for interindividual motor coordination and the joint Simon task in particular, apart from co-representation (e.g. Sebanz et al., 2003), action anticipation and adaptation (e.g. Keller et al., 2014), a joint focus of attention facilitating action monitoring (e.g. Tsai et al., 2006), and attribution of agency and goal-directed behavior (e.g. Stenzel et al., 2012) are relevant socio-cognitive skills to balance the integration and distinction between self- and other-generated actions.

4.2. Ontogenetic development, clinical studies, and implications for socio-cognitive requirements

To further evaluate to what extent co-representation in the joint Simon task may be a cognitively demanding high-level mechanism, relying on ToM reasoning, and/or advanced executive function skills, we now discuss developmental studies, and end the section by briefly reviewing clinical studies.

Children seem to understand and start to participate in simple forms of joint action based on shared intentionality from around 12–18 months even though young children’s engagement in joint action often requires scaffolding by an adult (Carpenter, 2009; Henderson and Woodward, 2011; Warneken and Tomasello, 2007). By age 2, children are capable to solve simple cooperation tasks with adults and peers (Brownell et al., 2006; Warneken et al., 2006). During the third year of life, they acquire an understanding of social conventions (social norms such as obligations and commitments towards the joint action partner, Gräfenhain et al., 2013, 2009; and rules of games, Rakocy et al., 2008), along with more sophisticated linguistic competences (Tomasello and Farrar, 1986). At the same time, they begin to engage more actively and explicitly in coordinated actions (Brownell et al., 2006), and by age 3, they are able to engage in more complex cooperation tasks involving complementary roles, and show higher accuracy and less variation in their timing for coordinating actions (Ashley and Tomasello, 1998; Meyer et al., 2010). 3.5–4-year-old children reliably coordinate actions with a partner during joint play and problem solving, show commitment by supporting each other’s actions with respect to shared goals, and flexibly adjust their actions to their joint action partner’s changing constraints (Hillmann et al., 2012; Kirschner and Tomasello, 2010; Meyer et al., 2016). Following the task needs, they coordinate their decision-making and actions with the partner by means of communicative cues (e.g. gestural or verbal cues) and visual monitoring (e.g. unidirectional or simultaneous bidirectional gaze at partner, or mutual eye contact), or teach less-skilled partners (Ashley and Tomasello, 1998;
The developmental trajectories during early childhood to engage in joint action with a peer are likely to relate to simultaneous developments in socio-cognitive skills such as perspective taking (Moll et al., 2013; Tomaszello and Hamann, 2012), action monitoring (Carpenter et al., 2005), and joint attention (Brownell et al., 2006). As such, by learning to simultaneously attend to the partner’s role, children become more proficient in action imitation (e.g. Fletcher et al., 2012), or by acquiring an understanding that anyone, including the self, perceives the world in any particular way among others, children may learn to not only take but also confront visual perspectives (Moll et al., 2013). The skills for inhibitory control (e.g. Carlson, 2005), and for ToM (e.g. Wellman et al., 2001) also typically follow a developmental progression, and likewise have been linked to improved joint action performance. For instance, higher accuracy in action coordination in an alternating turn-taking game was related to better inhibitory control skills in 2.5-year-olds (Meyer et al., 2015). Studies with school-aged children and young adults showed that more successful coordination of decision-making to choose a single response option out of several was related to increased ToM understanding (Curry and Chesters, 2012; Grüniesen et al., 2015), or that improved coordination in a joint ball-steering labyrinth game was related to increased emotion understanding (Viana et al., 2020).

Developmental research found a joint Simon effect in children between the ages of 4 and 5 years, but not in 2- and 3-year-olds (Milward et al., 2014; Saby et al., 2014). Thus, co-representation as measured with the joint Simon task appears rather late in ontogeny, which suggests that the onset of co-representation requires particular cognitive abilities only developing around that age, arguably linked to ToM understanding and inhibitory control skills. Alternatively, however, the lack of co-representation in children younger than 4 years may be an artifact of the specific task designs (Milward et al., 2017; Saby et al., 2014), which appear to play an important role (see paragraph 3.1.2). The only study that directly addressed the role of ToM in children between the ages of 4 and 5 years found that it was not a prerequisite (Milward et al., 2017). Rather, stronger ToM and inhibitory control abilities were associated with reduced interference from co-representing the partner in the joint Simon task, suggesting that ToM does not play a role in the emergence of co-representation, but is rather used to resolve conflict and correctly distinguish between self- and other-representation (i.e. SO distinction). The conclusion of these ontogenetic results is thus fully in line with the conclusion of the neurobiological studies summarized in 4.1.2 and 4.1.3.

These conclusions are also in line with studies focusing on socio-cognitive impairments, which provide rather indirect evidence for a link between the onset of co-representation and mature ToM understanding. Clinical studies have been conducted with participants who showed impairments of social cognition, due to either brain injuries, autism spectrum disorder, or schizophrenia. For instance, adult neurological patients with lesions either in the frontal or posterior parietal cortex (PPC) and the TPJ, and with impaired ToM (failing in first- and second-order ToM tasks reflecting both an individual’s knowledge of another’s beliefs, and the knowledge inferred about the beliefs of another person about a third party) did not show the joint Simon effect. Nonetheless, an effect emerged when the PPC/TPJ patients were explicitly instructed to attend to the co-actor (Humphreys and Bedford, 2011). In another study, adults with autism or Asperger Syndrome with impaired ToM (observed in worse performance in ToM tasks than control participants, however, all individuals except one passed either first- or second-order ToM tasks), showed a significant Simon effect (Schnitz et al., 2005b). Further, individuals with schizophrenia suffering from a disturbed sense of agency (often showing abnormal functioning in the MPC and TPJ) did not show the joint Simon effect (Liepelt et al., 2012). Thus, even if clinical studies similarly involve impaired ToM, details such as degree of impairment and exact location of affected brain areas might play an important role in explaining the outcome of the effect.

In sum, clinical studies are rather inconclusive, whereas current developmental studies suggest the emergence of co-representation in human children around age four, coinciding with the age around which a full-fledged, explicit ToM becomes functional (Wellman et al., 2001). Nevertheless, the role of ToM and inhibition in co-representation appears not to enable co-representation. Rather, stronger ToM and inhibitory control abilities were found to be linked to weaker co-representation in 4- to 5-year-old children (Milward et al., 2017). This view is consistent with studies on socio-cognitive impairment (Liepelt et al., 2012; Sebanz et al., 2005b) and sense of agency (Stenzel et al., 2014) suggesting attribution of goal-directed behavior (Tsai et al., 2008) and agency (Stenzel et al., 2012) to be important components for integrating own and others’ actions into a common representation (Sebanz et al., 2007). Findings from different studies with an emphasis on neurobiological, ontogenetic, and phylogenetic approaches (see below) thus suggest that full-fledged ToM reasoning may not be necessary for the onset of co-representation, and that co-representation may emerge from a rather implicit, automatic process. The lack of co-representation in children younger than 4 years may be linked to the task design and a lack of variables sensitive enough to detect co-representation (Milward et al., 2017; Saby et al., 2014), rather than a true negative result. In fact, when young children were tested with the identical prime version of the task, they showed evidence for co-representation from 2 to 3 years on (i.e. as soon as they would participate in the task), and its strength decreased rather than increased with advancing age (Miss, 2022). This suggests that children co-represent each other automatically from early on but become increasingly proficient at inhibiting co-representation in this task in which co-representation causes interference and reduces cooperation success, and thus again that co-representation in the joint Simon task perse is a low-level, automatic process.

4.3. Phylogeny and implications for socio-cognitive requirements

The proposal of co-representation as a social and cognitively non-demanding mechanism is further supported if it is also observed in other socially-living animals that may not be renowned for particularly strong general cognitive skills. Indeed, apart from humans, co-representation has been found in adult common marmoset monkeys (Miss and Burkart, 2018), Tonkean macaques, and brown capuchin monkeys (Miss et al., 2022a). Marmosets are cooperative breeders (i.e. other group members than parents significantly contribute to rearing offspring, Burkart et al., 2009; Hrdy, 2009). Group members are highly interdependent (de Oliveira Terceiro et al., 2021) and linked to shared infant care (Erb and Porter, 2017), marmosets routinely engage in cooperative behavior in their everyday life, as for instance food sharing (Guerreiro Martins et al., 2019), cooperative vocal turn-taking (Takahashi et al., 2013), or coordination of mutually exclusive activities among group members such as infant carrying vs. anti-predator behavior or feeding vs. vigilance (Brieger et al., in press; Snowdon, 2001). However, no evidence suggests they would have ToM abilities comparable to 4-year-old children, nor do they have particularly strong inhibitory control abilities (Burkart et al., 2017; Burkart and van Schaik, 2010). Tonkean macaques and capuchin monkeys, in contrast, are independent breeders and thus comparatively less cooperative in their daily life (Burkart et al., 2014, 2022; Janson, 1985; Mendres and de Waal, 2000; Petit et al., 1992; Thierry et al., 1994). They too show inhibitory control and ToM abilities that by far do not match 4-year-old humans (MacLean et al., 2014).

These recent findings of co-representation in primates suggest that co-representation shows a phylogenetic history that is shared with at least haplorrhines and was likely already present before the split in platyrhines and catarrhines (Miss et al., 2022a). The presence of co-representation across these different clades of primates suggests a general mechanism in primates, potentially activated automatically when they act together, given that dyads are tolerant enough to engage in a cooperative task at all (Miss et al., 2022a).

With regard to the strength of co-representation, the findings
indicate that it was strongest in the least cooperative species and the most cooperative one was better at suppressing co-representation, and thus showing SO distinction to improve joint performance. Thus, intriguingly, among the three primate species tested, the highly cooperative marmosets (Berkart et al., 2022) showed the highest flexibility to suppress co-representation in this task, and this led to the highest cooperation success. Therefore, the flexibility to regulate (and suppress if necessary) automatic co-representation might be enhanced in species in which individuals are routinely exposed to SO integration–distinction conflicts during joint actions from an early age on (Fig. 4). Moreover, the marmosets most frequently used pre-decision gaze at partner and were the only ones to also engage in mutual gaze (i.e. simultaneous bidirectional gaze at partner) in the joint task when coordinating their actions with a conspecific and facing ambiguity whose turn it was to perform the response action (Miss et al., 2022a). Partner-directed gaze may thus help to detect coordination affordances and to regulate the balance between SO integration and distinction, and facilitate cooperation in the joint Simon task. Comparatively weak co-representation and thus very high cooperation success, and frequent pre-decision (mutual) gaze at partner (as well as frequent gestural and verbal communicative cues targeting the response selection problem) was also observed in 2- to 4-year-old children when tested with the identical primate joint Simon task design (Miss, 2022). Even though we currently lack a directly comparative study in human adults tested with the primate task design involving a joint reward per successful trial, in human adults, social commitment and, thus, actor interdependence is likely established (though less overtly), for instance by giving an initial verbal instruction to perform the task together with the partner. Moreover, in human experiments, many other variables, such as signing up for, and committing to, the whole experiment, feelings of responsibility, or concern for reputation can serve as social incentives. In numerous studies in human adults, the observed joint Simon effect is often weak, observable only in marginal differences in response latencies varying mostly between 10 and 30 ms, and errors are commonly rare if occurring at all (e.g. Kiernan et al., 2012; Pfister et al., 2014; Sebanz et al., 2003). This may indicate that humans, having advanced cognitive abilities and at the same time a
hyper-cooperative lifestyle (Melis and Semmann, 2010) characterized by shared infant care among group members (Hrdy, 2009) as in mar- mosets, are most proficient in suppressing automatic co-representation when necessary (Fig. 4).

Therefore, intriguingly, these results in primates suggest that not necessarily species with the biggest brains and thus highest general cognitive ability (Burkart et al., 2017), but those more systematically relying on cooperation during their everyday life, particularly during offspring care (maromasets and humans) are better at flexibly suppressing co-representation in the joint Simon task (Miss et al., 2022a). Moreover, pre-decision communicative cues such as visual monitoring (unidirectional or bidirectional gaze at partner, gaze alternation, mutual eye contact, e.g. Duguid et al., 2014; Siposova et al., 2018; Wyman et al., 2013) and gestural or verbal cues (e.g. Duguid et al., 2020; Warneken et al., 2014) may serve to detect coordination affordances and help to resolve SO integration – distinction conflicts in dyadic action tasks (see also paragraph 4.2).

In sum, the comparative data from primates further support the view that co-representation is a rather automatic process not requiring advanced cognitive abilities (e.g. inhibitory control or ToM skills). Moreover, the process to prevent automatic co-representation from interfering with cooperation success appears to not primarily result from strong executive functions like general inhibitory control ability, but from the specific suppression of co-representation, which is likely learned during frequent exposure to SO integration – distinction conflicts in joint action contexts, and also involves social coordination smoothers such as mutual gaze.

5. Discussion

5.1. Co-representation as a social and automatic mechanism

Joint action permeates our everyday life and has even been argued to have played a key role for the evolution of our minds. It can vary from behavior that is driven by simple rules or spontaneous reflexes (emergent coordination), to behavior that is adjusted to interaction partners in pursuit of a joint outcome (goal-directed coordination), to relying on complex mind reading and meshing intentions as in shared intentionality (e.g. Bratman, 1992; Knoblich et al., 2011; Tomasello et al., 2005, Fig. 1). The conceptual fuzziness of joint action and related terms, and the ensuing difficulties to operationalize it in a straightforward way makes it challenging to evaluate if and how these phenomena relate to each other, and how they evolved. One of the key questions is whether engaging in joint actions is a cognitively demanding behavior or not, which may be the case if it is based on co-representation and shared intentionality.

In this paper, we chose a paradigm-driven, bottom-up approach to contribute to integrating joint action-related research, focusing on co-representation as measured with joint Simon tasks (e.g. Sebanz et al., 2003). The fact that the strength of the joint Simon effect co-varies with social factors in humans when strong contrasts exist (e.g. ingroup vs. outgroup co-actor, McClung et al., 2013), and the finding of coupled neural oscillations in co-actors’ brains when jointly engaged in action coordination (e.g. Djalovski et al., 2021; Hasson et al., 2012) both suggest that the effect is genuinely social. As highlighted in Table 1, this is in particular the case when experimental designs make the joint goal salient, for instance by rewarding both partners for a correct answer by either of them (e.g. Miss and Burkart, 2018).

The findings reviewed from the fields of neurobiology, developmental, and evolutionary biology suggest that co-representation is not a complex mechanism relying on strong cognitive skills, but rather a basic process that emerges automatically and implicitly (see also Sebanz and Knoblich, 2021; Southgate, 2020). First, co-representation was observed in human adults despite its detrimental effect on joint task performance in the joint Simon task (e.g. Sebanz et al., 2003; Tsai et al., 2008). Second, neurobiological findings are most consistent with an automatic co-representation mechanism. If it interferes with cooperation success, it needs to be suppressed. Intriguingly, increasing evidence suggests some continuity and neurobiological overlap of simple action co-representation and higher-level co-representation, for instance of perspectives, emotions, or beliefs. Third, ToM seems not a necessary requirement for the onset of co-representation during development, but may rather be involved in SO distinction in children, leading to weaker co-representation in the joint Simon task (Milward et al., 2017). Finally, co-representation was found as well in various monkey species (Miss et al., 2022a; Miss and Burkart, 2018), who have limited inhibitory control and ToM abilities compared to humans. Altogether, this suggests that co-representation as measured with joint Simon tasks is fundamentally social and the result of an automatic mechanism.

5.2. To merge or not? – task designs can make self-other integration helpful or hindering

Task setup and social context can temporarily affect individuals’ representations in a way that they perceive themselves as either part of a social context (as interdependent) or more in isolation (as independent). Accordingly, if shared features (as opposed to discriminating features) are more salient, this likely increases the perceived overlap between oneself and other (Markus and Kitayama, 1991). For instance, participants who experienced interdependence priming (by having to circle interdependent pronouns such as ‘we’, ‘ours’ in an assay) between experimental blocks of the joint Simon task showed stronger co-representation than participants who experienced independence priming (by having to circle independent pronouns such as ‘I’, ‘mine’ in the assay, Colzato et al., 2012). Therefore, as discussed earlier, joint Simon task designs generating interdependence, rather than independence, seem to make it more likely that one’s own and the other’s actions are integrated in a common representation, which can be modulated in different ways (see paragraph 3.1.2). For instance, introducing a salient joint goal with a joint reward structure (Miss and Burkart, 2018) or with explicit verbal instructions to cooperate rather than to compete (e.g. Iani et al., 2011) likely emphasizes interdependence. Other examples are explicit instructions to take the co-actor’s perspective (Müller et al., 2011a, 2011b), or the introduction of a supportive (friendly vs. antagonistic, Hommel et al., 2009), socially close (a friend or spouse vs. a stranger, Ford and Aberdeen, 2015; Shafaei et al., 2020), or highly familiar (ingroup vs. outgroup, McClung et al., 2013) co-actor.

Thus, the task design needs to be selected carefully when studying action coordination and co-representation. An important distinction is that in some set-ups (e.g. synchronization tasks), cooperation success increases with SO integration and co-representation, whereas in others, such as the joint Simon task, cooperation success crucially requires SO distinction. In the joint Simon task, automatic co-representation may thus increase with social closeness, but individuals may counteract it at the same time by inhibiting its expression because this maximizes cooperation success. The strength of co-representation measured will thus be the result of these two processes, making it difficult or impossible to disentangle the contribution of each of them, and how they are influenced by social factors. Moreover, when investigating the ontogenetic onset of co-representation, observations will likely differ depending on whether the measure is the joint Simon task or another joint action task that might only involve co-representation but not require inhibition. In the future, it would therefore be crucial to develop paradigms that allow quantifying both processes separately, i.e. the tendency to merge as well as the ability to enhance SO distinction.

5.3. Cooperative flexibility mediated by social cues

When co-representation hinders cooperation success, it needs to be suppressed. This appears to involve a specific suppression mechanism of SO integration (Hamilton, 2021; Steinbeis, 2016), rather than general executive functions such as inhibitory control ability (Santiesteban
Among the monkey species tested with a joint Simon task, common marmosets were found to be best at flexibly regulating the strength of co-representation, and to most frequently use mutual gaze as pre-decision communicative cue (Miss et al., 2022a). Preschool children tested with the identical primate joint Simon task design rarely made errors in manual response choices, resulting in very high cooperation success. Nonetheless, a joint Simon effect indicative of co-representation was observed in their first orienting direction, which is a more subtle behavior and thus likely provides a more sensitive outcome variable. Interference effects were stronger in the younger than the older children. Moreover, when determining whose turn it was to respond, younger children frequently used pre-decision visual monitoring of the partner and communicative (verbal and gestural) cues referring to the response selection problem. Thus, the younger children, who had more difficulties to suppress co-representation, were more likely to use pre-decision social cues, likely to receive social information and support to resolve the conflict between SO integration and distinction (Miss, 2022). Thus, the ability to co-represent a partner’s actions during joint actions is not what makes humans unique compared to other primates. Humans have particularly large, powerful brains, which permits the development of exceptionally strong ToM and language skills. However, differences appear to arise earlier in ontogeny already before the acquisition of strong cognitive abilities, suggesting a link with motivational rather than cognitive factors, reflected for instance in stronger preferences for joint engagement in activities, sharing, and helping (Bullinger et al., 2011; Grafenhain et al., 2009; Kirschner and Tomasello, 2010; Rekers et al., 2011; Tomasello et al., 2005).

When routine cooperation is prevalent in a species, such as in humans or marmosets, individuals have recurrent opportunities to gain experience in joint activities and to become competent cooperators, for instance in cooperative problem solving (e.g. Martin et al., 2021). Such joint endeavors include for example cooperative infant care-taking, food sharing, cooperative vocal communication, or the coordination of mutually exclusive activities among group members (Burkart et al., 2022). This may facilitate the acquisition of enhanced abilities in social learning, such as behavior copying and imitation (Fletcher et al., 2012; in marmosets, Voelkl and Huber, 2007, 2000), communication (Goldstein and Schwade, 2008; in marmosets, Gultekin and Hage, 2017; Takahashi et al., 2017), coordination of attention and action (Bakeman and Adamson, 1984; Moll et al., 2008), or understanding and confronting visual perspectives (Moll et al., 2013). Thus, during development, through interactions with their mothers, but also other caregivers and peers, young infants may learn and practice continuously when to merge and when to dissociate themselves from the other, while general cognitive skills (executive functions such as inhibitory control, and ToM) develop in parallel and may, at an advanced stage, come to support this process. Experience-based cooperative flexibility may include the learning of coordination smoother (e.g. mutual gaze) when coordination affordances are high (e.g. Duguid et al., 2014; Warnken et al., 2014). In music ensemble performance for instance, visual monitoring aids coordination during temporal instability (Bishop et al., 2019) and in natural conversations, mutual eye contact can serve as a signal of shared attention and may aid to synchronize coordinations (Woolley and Wheatley, 2021). A cooperatively breeding bird species, the Arabian babblers, was observed to use gaze alternation to co-orientate attention during communicative signaling for coordinating joint travel (Ben Mocha et al., 2019). In a dyadic cooperative pulling task, brown capuchin monkeys monitored each other’s actions and were less successful when visual contact between them was blocked (Mendres and de Waal, 2000). Therefore, a systematic study of the role of gaze behavior and (non)verbal cooperative cues in action coordination in humans and animals (see also Duguid and Melis, 2020; Heesen et al., 2021; Iki and Hasegawa, 2021) may be a promising subject for further investigations.

In conclusion, the current evidence most likely suggests that a lack of co-representation in humans, sometimes observed in adults but also in young children in joint Simon tasks is not because co-representation is cognitively too demanding. Rather, humans appear to have high proficiency in suppressing co-representation whenever necessary, which is supported by pre-decision communicative cues that can help detecting coordination affordances and regulating the balance between SO integration and distinction.

5.4. Conclusions and future directions

Our paradigm-driven bottom-up approach of aligning findings on co-representation across fields and integrating them in the broad context of joint action research turned out to be a very fruitful first step but is clearly not the end point. With a primary focus on the joint Simon task, results across fields suggest co-representation as a social and rather automatic mechanism, not relying on advanced cognitive abilities. This approach further provided insights into how successful dyadic action coordination can be achieved. Namely, besides the involvement of strong general cognitive skills like inhibitory control and ToM, SO-conflict resolution may also be achieved by engaging in cognitively non-demanding social interaction dynamics like mutual gaze and verbal or gestural cues when coordinating actions with a partner. The proposal that cooperative flexibility can be acquired over time can in principle be tested with training studies with primates (rather than humans who most likely show ceiling effects even early during ontogeny). Testing more groups and more species will be needed to further consolidate and strengthen these proposals, as well as to increase our knowledge on how SO integration and distinction of motor actions are linked to SO integration and distinction at higher-level representations of minds involving for instance perspectives, emotions, or beliefs. Whether co-representation extends beyond dyads and is involved in coordinating group behaviors is another interesting question for future studies. Further, to get rid of the need of suppressing co-representation to improve joint performance and be successful in the joint Simon task, it appears helpful to experimentally disentangle the processes of SO integration and distinction in joint actions, and assess subjects either in tasks in which SO integration and co-representation is favorable for cooperation success, or in tasks that critically require SO distinction for successful joint performance.

Further, a similar approach of analyzing and integrating findings across fields can be used for other critical components of joint action. For instance, our focus here was predominantly on cognitive factors, whereas we mostly did not address the motivational factors important for joint action and shared intentionality. A critical motivational factor to engage in joint action and eventually share mental states with others may be proactive prosociality for which we can take an analogous approach. Like co-representation, proactive prosociality does not seem to be a cognitively highly demanding trait and likewise be particularly strong in species with a highly cooperative lifestyle (Burkart and van Schaik, 2020; Hrdy and Burkart, 2020).

Other recent approaches are developing innovative operationalization criteria to get a firmer empirical grasp on joint action. Recently, joint action has been construed as a process rather than as a-result, which makes it possible to empirically scrutinize whether opening and closing phases have behavioral, quantifiable markers of joint commitment (Genty et al., 2020; Heesen et al., 2020, 2017).

From a neurobiological perspective, hyperscanning results that reveal brain-to-brain coupling during real-time joint action look promising for future research. In particular, an investigation of the role of co-representation at higher conceptual levels involving representations of minds has the potential to move the field forward. Exact functions of brain areas and neural rhythms involved in joint actions and how they interplay is thus a topic of ongoing research. Importantly, conclusions...
drawn from studies of neural mechanisms need to consider the task design, since we do not know if behaviors in solitary tasks requesting either one’s own action performance or pure observation of others’ actions rely (partly) on the same or different mechanisms as behaviors in cooperative tasks including shared goals, mutual adaptation, and dynamic social exchanges. Moreover, neuroimaging techniques are often limited in how accurately they can delineate the underlying neural computations (Lamm et al., 2016). Therefore, it is important to use a combination of methods generating converging evidence for the same type of conclusions (Hamilton, 2021; Lamm et al., 2016). We fully agree with Krakauer et al. (2017) and Sebanz and Knoblich (2009) that behavioral work should first provide a basic understanding of what a task measures and only afterwards be complemented by the study of neural component processes for testing causality.

Joint action and the different conceptualizations of this phenomenon in different fields makes it a particularly complex topic. The use of a variety of experimental and theoretical approaches is therefore fully warranted. Nevertheless, the ultimate goal should be to bring the results of these diverse approaches together and to integrate them in a meaningful way. Our contribution can be seen as one more step in this direction.

Funding

This work was supported by the Swiss National Science Foundation, Switzerland (grant number 31003A_172979), the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No. 10010295), the NCCR Evolving Language, Swiss National Science Foundation (agreement number 51NF40_180888), and by the A.H. Schultz Foundation, Switzerland. The funders had no role in the design, decision to publish, or number 51NF40_180888), and by the A.H. Schultz Foundation, Switzerland.

Declaration of Competing Interest

None.

Acknowledgments

We would like to thank Carel van Schaik for valuable input, and Pierre-Henri Moreau for insightful discussions. We also thank the editor and two anonymous reviewers for their constructive feedback on previous versions of the manuscript.

References

Adriaanse, J.E.C., Koski, S.E., Huber, L., Lamm, C., 2020. Challenges in the comparative study of empathy and related phenomena in animals. Neurosci. Biobehav. Rev. 112, 62–82. https://doi.org/10.1016/j.neubiorev.2020.01.021.

Amoedo, D.M., Frith, C.D., 2006. Meeting of minds: The medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7, 268–277. https://doi.org/10.1038/nrn1894.

Aron, A., Aron, E.N., Smollan, D., 1999. Inclusion of other in the self scale and the structure of interpersonal closeness. J. Pers. Soc. Psychol. 63, 596–612. https://doi.org/10.1037/0022-3514.63.4.596.

Ashley, J., Tomasello, M., 2005. Role reversal imitation and language in mother-infant and peer-infant interaction. Child Dev. 55, 1278–1289. https://doi.org/10.1111/1467-9507.00059.

Atmaca, S., Sebanz, N., Prinz, W., Knoblich, G., 2008. Action co-representation: The joint SNARC effect. Soc. Neurosci. 3, 410–420. https://doi.org/10.1080/1747091080900968.

Atmaca, S., Sebanz, N., Knoblich, G., 2011. The joint flanker effect: sharing tasks with real and imagined co-actors. Exp. Brain Res. 211, 371–385. https://doi.org/10.1007/s00221-011-2703-9.

Baker, F., 2004. Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Dev. 55, 1278–1289. https://doi.org/10.1111/1467-9507.00059.

Ballon, L., Adamson, L.B., 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Dev. 55, 1278–1289. https://doi.org/10.1111/1467-9507.00059.

Ballon, L.B., Adamson, L., 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Dev. 55, 1278–1289. https://doi.org/10.1111/1467-9507.00059.

Bakeman, R., Adamson, L.B., 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Dev. 55, 1278–1289. https://doi.org/10.1111/1467-9507.00059.

Bakeman, R., Adamson, L.B., 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Dev. 55, 1278–1289. https://doi.org/10.1111/1467-9507.00059.

Bakeman, R., Adamson, L.B., 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Dev. 55, 1278–1289. https://doi.org/10.1111/1467-9507.00059.

Bakeman, R., Adamson, L.B., 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Dev. 55, 1278–1289. https://doi.org/10.1111/1467-9507.00059.
Duguid, S., Wyman, E., Bullinger, A.F., Herfurth-Majstorovic, K., Tomasello, M., 2014. How animals collaborate: Underlying proximate mechanisms. Wiley Interdiscip. Rev. Cogn. Sci. 11, 1–21.

Dittrich, K., Dolk, T., Rothe-Wulf, A., Klauer, K.C., Prinz, W., 2013. Keys and seats: Spatial response coding underlying the joint spatial compatibility effect. Atten. Percept. Psychophys. 75, 1725–1736. https://doi.org/10.3758/s13420-013-0524-z.

Djolavski, A., Dumas, G., Kinreich, S., Feldman, R., 2021. Human attachments shape empathy and friendship. Front. Psychol. 6, 962. https://doi.org/10.3389/fpsyg.2015.00962.

Dittrich, K., Krause, M.I., Rushton, G.D., Frank, N.R., 2002. Collective memory and spatial listing in animal groups. J. Theor. Biol. 218, 1–11. https://doi.org/10.1006/jtbi.2002.3065.

Cook, R., Bird, G., Catmur, C., Press, C., Heyes, C., 2014. Mirror neurons: From origin to function. Behav. Brain Sci. 37, 177–192. https://doi.org/10.1017/S0140525X13000903.

Cook, M.O., Kruzick, R., Rushton, G.D., Frank, N.R., 2002. Collective memory and spatial listing in animal groups. J. Theor. Biol. 218, 1–11. https://doi.org/10.1006/jtbi.2002.3065.

Cook, A., Bird, G., Catmur, C., Press, C., Heyes, C., 2014. Mirror neurons: From origin to function. Behav. Brain Sci. 37, 177–192. https://doi.org/10.1017/S0140525X13000903.

Cook, A., Bird, G., Catmur, C., Press, C., Heyes, C., 2014. Mirror neurons: From origin to function. Behav. Brain Sci. 37, 177–192. https://doi.org/10.1017/S0140525X13000903.

Cook, R., Bird, G., Catmur, C., Press, C., Heyes, C., 2014. Mirror neurons: From origin to function. Behav. Brain Sci. 37, 177–192. https://doi.org/10.1017/S0140525X13000903.

Cook, A., Bird, G., Catmur, C., Press, C., Heyes, C., 2014. Mirror neurons: From origin to function. Behav. Brain Sci. 37, 177–192. https://doi.org/10.1017/S0140525X13000903.
