Joint action and joint attention: Drawing parallels between the literatures

Sophie J. Milward [1,2,a] & Malinda Carpenter[3]

[1] Max Planck Institute for Evolutionary Anthropology, Department of Developmental and

Comparative Psychology, Leipzig, Germany

[2] University of Portsmouth, Department of Psychology, Portsmouth, UK

[3] University of St Andrews, School of Psychology and Neuroscience, St Andrews, UK

[a] Corresponding author: Email sophie.milward@port.ac.uk

Abstract

Two of the most important milestones in children's development are joint action (acting with others) and joint attention (attending with others). These are popular fields in both psychology and philosophy, but have formed surprisingly independent literatures despite the close similarities they share in terms of theoretical and methodological issues. This article systematically compares these fields and draws attention to specific and more general ways in which each could benefit from the other if communication between them were increased. We highlight a clear opportunity within these fields, but this could be a useful approach in cognitive science more generally.

Social interaction is a crucial and pervasive component of human behaviour, and is

consequently a skill at which we are highly adept. As a result, cognitive development research

has focussed heavily on the developmental milestones that allow us to perform both simple and

more complex joint behaviours. Two of the most important of these milestones are joint action

and joint attention. These abilities have attracted a huge amount of interest from both

philosophers and psychologists, studying both children and adults, but have each become the

centre of largely independent literatures. Despite the relative lack of communication between

these literatures, they have evolved in very similar ways and stumbled across similar issues. We

argue that they would benefit from direct comparison, particularly where each area has identified

different solutions that could be shared. Here we thus aim to initiate the building of a bridge

between the fields of joint action and joint attention from both a psychological and a

philosophical perspective. We start by highlighting the similar problems these fields have come

across when attempting to define these concepts, particularly when defining what level of

interaction constitutes "jointness." We go on to discuss how researchers in each field have

attempted to identify jointness using experimental studies, pointing out where cross-over could

be beneficial. This is followed by a discussion of how we achieve jointness when acting and/or

attending together. We then highlight some additional issues that can be compared across fields,

and provide ways in which comparison can aid progress in both methodology and theory.

Difficulties with definition

Joint action refers to the co-ordination of actions between two or more individuals. There

are different definitions of joint action that stipulate lower vs. higher levels of processing of the

other's mental states. For example, Sebanz, Bekkering and Knoblich (2006) use an inclusive

definition, proposing that joint action is "any form of social interaction whereby two or more

individuals coordinate their actions in space and time to bring about a change in the environment" (p. 70). This could include coordinated acts such as birds flocking, which do not necessarily involve understanding of the intentions of the other individuals in the group, but instead can be achieved by reactions to perceptual information from the other group members (e.g. direction of flight). At the other end of the scale, Bratman (1992) defines a specific type of joint action, Shared Cooperative Activity, as requiring both actors to understand the intentions of the other to achieve a shared goal, to be committed to the joint activity, to have common knowledge that this is the case and to be committed to supporting the partner. This is clearly a much more cognitively demanding definition involving bidirectional understanding of mental states and obligations.

Similarly, definitions of joint attention also range from very inclusive to very restrictive. For example, Butterworth (1995) simply defined joint attention as "looking where others are looking" (p. 29). This would include behaviours of non-human animals where both parties are attending to the same thing, but there is not necessarily any common knowledge between the two that they are doing so. In contrast, Tomasello (1995) specifies a more restrictive definition in which it is necessary that the individuals not only attend to the same object, but also both know together that they are attending to the object and each other's attention. This maps directly onto the common knowledge requirement in Bratman's (1992) definition of joint action.

It should be noted that for both joint action and joint attention it is not the case that one definition is correct and others are incorrect, in the sense that there is in fact a range of different levels of 'joint' behaviour. For example, within joint attention there are many other definitions that lie between the most inclusive and most restrictive (e.g. see Bakeman & Adamson, 1984; Leavens & Racine, 2009). One possibility suggested by Emery (2000) is to split these types of

behaviours into two categories: "joint attention", which includes behaviours without shared knowledge, versus "shared attention", which includes behaviours with shared knowledge. It would already be extremely helpful if researchers would use the terms "joint" and "shared" in this way universally in both fields. However, this distinction still may not be thorough enough to eradicate confusion entirely where further ranges of behaviours exist within each category (see Siposova & Carpenter, forthcoming, for a more detailed typology of different levels of joint attention and joint action). In short, it would be helpful if there was agreement both within and across fields about terminology so that the same terms are not used to describe different ranges of behaviour. This variation in definitions causes confusion in the literature (e.g. see Leavens & Racine, 2009, versus Carpenter & Call, 2013: each argue for or against joint attention in great apes, respectively, but use different definitions of joint attention). This variation in definitions also leads to some of the issues outlined below, such as how to fulfil a criterion of true jointness. Interestingly, the qualities about which scholars disagree are highly comparable across both literatures, for example with regard to the criterion of common knowledge. Thus, an important first step for both fields is to establish agreement on how to define the different levels of each phenomenon. Once this has been achieved, it will be easier to identify further parallels between the two literatures that can aid progress in both.

How can we identify jointness?

As individuals, it may seem trivially clear to us when we are involved in joint action and/or joint attention with another person (Reddy, 2003). However, this may not always be so easy to distinguish as an empirical observer. Here we outline two ways that researchers have come up with to overcome this problem: exploiting unwanted side-effects of jointness, such as self-other interference, and experimentally manipulating factors that might increase jointness and

measuring the corresponding changes in behaviour. Both of these approaches enable us to test the criteria laid out by the different definitions outlined above, to see whether fulfilling a given criterion produces changes in behaviour. If so, then the criterion can be assumed to be meaningful to jointness in some way.

Self-Other Interference. One particularly interesting phenomenon that is frequently cited as evidence of jointness is self-other integration and interference. This is the phenomenon that when individuals are participating in joint action, they often experience a merging of representations of self and other. This has been identified in several different fields, including joint task representations (Sebanz, Knoblich & Prinz, 2003), joint focus of attention (Böckler, Knoblich & Sebanz, 2012), perspective-taking (Surtees & Apperly, 2012), motor representations (Brass, Ruby & Spengler, 2009) and emotion understanding (Steinbeis, 2016). Influence between self and other can occur bidirectionally, so that an individual may either egocentrically project representations of the self onto another person, or alternatively find their own self representations influenced by representations of the other. This mechanism has been argued to aid us in predicting a coordination partner's actions and adapting our own accordingly (Vesper et al., 2010). Interestingly, it can actually result in reduced performance in some tasks. For example, individuals are less able to judge their own visual perspective on a scene if another individual is present who holds a different perspective (Samson, Apperly, Braithwaite & Andrews, 2010). This kind of interference from another person has been exploited in empirical research to demonstrate the presence of self-other influence, as well as to demonstrate its automaticity.

The phenomenon of self-other interference is evidence for a degree of jointness, and as such it is useful to study across development to see when children can be considered to be acting

jointly. Egocentrism in children is a well-established phenomenon, whereby individuals assume that their own thoughts and emotions are shared by others. This has been established in the fields of visual perspective-taking (Piaget & Inhelder, 1956) and mental state attribution (Wimmer & Perner, 1983), with children younger than around 3-4 years tending to respond as though others have the same perspective or mental state as they do even when this is not the case.

However, egocentrism (influence from self to other) is less indicative of jointness than influence in the opposite direction, because egocentrism does not necessarily require representation of the other person at all. In contrast, interference from other to self is direct evidence that the other is represented, and that this representation influences the pre-existing representation of one's own perspective or mental state. This has been researched to a lesser degree than egocentrism, but there is some evidence that children do experience such other-self interference. In the field of joint action, several studies have shown that when 4- to 5-year-old children perform a task alongside a partner, even when it is not necessary to represent their partner's different task, they make mistakes or hesitations consistent with doing so (Milward, Kita & Apperly, 2014; Milward, Kita & Apperly, 2016; Saby, Bouquet & Marshall, 2014). In joint attention research, adults have been shown to experience similar interference from another person's focus of attention (Böckler et al., 2012). No work has yet been done on this with children in a joint attention context per se, but similar results have been found in a perspective-taking context. Six-year-olds respond more slowly when judging their own perspective on an object presented on a computer screen if an avatar on the screen had a different perspective (Surtees, Butterfill & Apperly, 2012). It would be interesting to adapt the Böckler et al. (2012)

task to investigate whether similar automatic self-other interference can be found with children in a joint attention context.

Although self-other interference is discussed as evidence for jointness, it also occurs even in circumstances in which the individuals involved are acting or attending merely in parallel to one another, not truly jointly. For example, in the joint action studies described above, the two participants in the task are each required to follow their own task rule. Although they are aware of their partner's task rule, it is completely irrelevant to and independent of their own rule. Likewise, in the joint attention and perspective-taking studies above (Böckler et al., 2012; Surtees et al., 2012), the attention or perspective of the partner is completely irrelevant to the participant, and yet participants still represent it. This is interesting and informative, as it suggests that this is an automatic process, which may have evolved as a result of humans' highly collaborative nature (Tomasello, Carpenter, Call, Behne & Moll, 2005), but which is over-active even in circumstances in which it does not provide the benefit of aiding prediction of a collaborative partner. It can thus be triggered merely by the observation of or belief that another person is performing a certain task role, regardless of whether that role is being performed as part of a collaborative effort. This thus limits the extent to which this phenomenon can inform us about different levels of joint action and attention. It would be informative to investigate whether even stronger co-representation effects are seen in higher level scenarios involving common knowledge.

Manipulating the joint context. Another approach to identifying jointness is to experimentally manipulate factors hypothesised to elicit jointness and identify some of the observable consequences of these manipulations. For example, researchers often compare the effects of scenarios in which individuals act or attend in synchrony or congruence with one

another (fulfilling lower level definitions of jointness) to those in which they act in asynchrony or incongruence. At a higher level, researchers also often compare situations in which individuals share goals or attention with one another (fulfilling higher level definitions) to scenarios in which they either act in parallel or observe others acting individually.

In joint action, researchers have manipulated low level synchrony in order to observe how this affects participants' attitudes and behaviour to another person. For example, Hove and Risen (2009) found that manipulating individuals' finger-tapping synchrony with an experimenter led to modulation of their ratings of affiliation, with higher affiliation to synchronous experimenters (see also, e.g. Cirelli, Einarson & Trainor, 2014; Tuncgenc, Cohen, & Fawcett, 2015, for similar results with infants). At a higher level, studies have shown differences in behaviour depending on whether participants were collaborating on a joint goal or not. For example, Hamann, Warneken, Greenberg and Tomasello (2011) found that 3-year-old children, but not chimpanzees, shared resources with their partner more equally when the resources were acquired as a result of collaboration rather than a windfall or parallel work. Additionally, sharing a joint goal can induce a feeling of commitment or obligation, which may result in persisting at an activity for longer than if no obligation is present. Hamann, Warneken and Tomasello (2012) found that 3.5-year-olds were more likely to continue an activity even after they had already received their reward but their partner had not yet, when the activity was collaborative than when it was non-collaborative. Further, Gräfenhain, Carpenter and Tomasello (2013) found that 3-year-olds were more likely to wait for and help their partner if they had made a verbal joint commitment than if they had not. Finally, Plötner, Over, Carpenter and Tomasello (2015) showed that 5-year-old children were more likely to trust, help and like their collaborative partners over individuals they had not collaborated with, even after the

collaborative activity was over, suggesting that jointness alters the attitudes that individuals have towards others. These studies show how sharing a collaborative experience can alter the behaviours and attitudes of individuals, as compared to individual experiences involving identical actions.

Similarly, studies on joint attention have shown effects of gaze following (i.e. lower level joint attention) on attitudes towards gaze partners and processing of observed objects. Research with adults found that participants prefer avatars who followed their gaze to an object over those who looked away (Bayliss & Tipper, 1996) and they also prefer objects that were gazed at mutually over those that were not (Bayliss, Paul, Cannon, & Tipper, 2006). Additionally, infants show a novelty preference for objects that they have not previously seen an adult gazing toward (Reid & Striano, 2005; see also Reid, Striano, Kaufman & Johnson, 2004 for differences in ERP responses). At a higher level of joint attention involving shared experiences or common ground, differences have been shown in infants' communication depending on whether they had just shared attention to an object or event with their partner or not. These studies have found that 1-year-old infants both initiate gestural communication (Liebal, Carpenter & Tomasello, 2010) and respond to others' gestural and verbal communication (Ganea & Saylor, 2007; Liebal, Behne, Carpenter & Tomasello, 2009; Saylor & Ganea, 2007) differently depending on what objects or events they had previously attended to jointly with their partner (vs. another partner). For example, the infants in the Liebal et al. (2010) study pointed to objects that were most relevant to the experience they had just shared with their communicative partner. With a series of control conditions, Moll, Richter, Carpenter and Tomasello (2008) demonstrated that children's responses were based on particular experiences they had previously shared with the particular partner with whom they were interacting, rather than on their own individual experiences, or

their partner's own individual experiences. Furthermore, Moll, Carpenter and Tomasello (2007) found that 14-month-old infants could only correctly identify which objects an experimenter had previously seen when they had previously shared attention to those objects with her – they could not do this when they had previously merely observed the experimenter attending to the objects herself. Although the same information regarding what the experimenter had seen was available in both scenarios, the addition of sharing looks between the child and the experimenter provided important extra information for infants at this age. This is a clear demonstration of the distinctive value of jointness for processing of information from social interactions.

The fields of joint action and joint attention have thus developed a range of methods for measuring jointness. Within joint action, researchers have investigated interference from automatic co-representation and changes in prosocial behaviour, attitudes towards and commitment to a partner following joint action, whereas within joint attention researchers have focussed on differences in preferences, processing and communication following joint versus individual experiences. It might be valuable to introduce these respective methods to the other field, for example by investigating whether there are differing degrees of self-other interference in high versus lower level joint action and joint attention, or differences in communication in joint vs. parallel action contexts (see below for a step in this direction). Again, researchers should keep in mind which level of jointness their studies are demonstrating and adopt appropriate terminology accordingly.

How do we achieve jointness?

Having established that whether interactions are joint or not seems to result in important differences in behaviour and attitudes, the question remains how we coordinate or share action and attention with others in the first place. One answer to this question for joint action is low

level action coordination. For example, individuals in close proximity often end up synchronising their rhythmic movements, such as walking in step with one another (Van Ulzen, Lamoth, Daffertshofer, Semin & Beek, 2008). This is just one of several lower level mechanisms (see also the section on *top-down versus bottom-up processing* below) that seem to create a level of jointness at an implicit level. However, this cannot explain higher levels of jointness, where not only do I coordinate with you at an implicit level, but we both intend to collaborate towards a goal that we know is shared between us. Similarly, low level mechanisms can go some way toward explaining joint attention. For example, not only do we often automatically follow others' gaze, we also are sensitive to our own gaze being followed (Edwards, Stephenson, Dalmaso & Bayliss, 2015). We thus have very effective mechanisms for knowing when we are looking at the same thing as someone else. However, this does not explain how we achieve higher levels of jointness, where not only do I know that you have seen what I have seen and vice versa, but we also *know this together*.

One way in which we might go a step further in achieving a higher level of jointness is via communication. This was first suggested in the developmental joint attention literature. The idea is that true sharing of attention – knowing together that we are sharing attention – is achieved via communication (Carpenter & Liebal, 2011). This can be verbal or gestural communication once children have mastered these, for example when an infant points to an object for her mother and the mother responds with a positive comment about it. However, one even simpler way of communicating *knowing together* is simply through a "sharing look," when two individuals alternate between looking at the object of attention and engaging in communicative eye contact with each other (Carpenter & Liebal, 2011). Infants begin engaging in these communicative, "sharing looks" from around 9 months of age (e.g., Venezia, Messinger,

Thorp & Mundy, 2004), suggesting that they may be capable of a high level of jointness from very early in ontogeny.

The importance of sharing looks has also been investigated in the developmental joint action literature, somewhat later in development. Wyman, Rakoczy and Tomasello (2012) found that these types of looks can facilitate joint action by helping partners solve coordination problems. They had 4-year-old children play a coordination game in which children needed to decide whether to risk playing cooperatively with their partner for a large reward, or play individually for a smaller reward. They found that children were more likely to choose to coordinate for the larger reward when their partner made eye contact and smiled than when she did not, suggesting that they interpreted this sharing look as a signal for joint action. Following up on this finding, Siposova, Tomasello and Carpenter (submitted) showed that (silent) communicative, but not non-communicative, eye contact can even be taken by 6- to 7-year-old children as a commitment on their partner's part to cooperate, as evidenced by children's normative protests (e.g. "You should do what you said you would do!") when their partner subsequently chose to play individually.

While developmental researchers have investigated the role of communication in establishing joint action, researchers studying adults have focussed on its role in maintaining joint action once it is established. They have documented a form of sensorimotor communication in joint action, in which individuals modulate the actions that they perform as part of the task in order to make them more predictable to their co-actors (Vesper et al., 2017). For example, Goebl and Palmer (2009) found that pianists playing a duet raised their fingers higher when auditory feedback between the two actors was reduced, suggesting that they modulated their actions in order to aid predictability and coordination. This has yet to be studied

in children and might provide another useful indicator of jointness, as sharing looks have done for both joint action and joint attention.

Communication, both verbal and non-verbal, has been highlighted as a way in which a higher level of jointness may arise. This has been discussed independently in both the fields of joint action and joint attention, but both fields could benefit from sharing ideas. For example, one step could be to investigate the role of communication not just in creating jointness in joint attention, but also in maintaining jointness, as has been studied in the adult joint action literature.

Issues beyond jointness

Content of representations. Beyond the issue of how much is shared between participants in an interaction, a further issue that arises from questions surrounding social representations is what the actual content of these representations is, regardless of the degree to which they are shared. This is a question that has been raised independently in the joint action and joint attention literatures, yet each has information that they can contribute to the other.

In joint action research, this question has arisen in the debate around task co-representation. Studies showing interference from a partner's task do not necessarily demonstrate which aspect of the task is being represented. One possibility is that it is the partner's action that is represented, which is supported by research showing interference when performing an action and observing another's incompatible action (e.g. Kilner, Paulignan & Blakemore, 2003). This should mean that an identical action performed on a different object (for example me tapping a green button and you tapping a blue button) would not cause interference. However, in most studies showing co-representation interference, the action is, in fact, the same for both individuals, but is directed towards different objects (e.g. different response buttons), suggesting that it cannot be merely representation of the action that causes interference. This led

researchers to argue that what is being represented is the task rule, that is, to make a certain response to a certain stimulus (Sebanz et al., 2003). However, a study by Wenke, Atmaca, Holländer, Liepelt, Baess and Prinz (2011) found evidence that partners in a joint task do not represent the exact stimulus-response mappings of their partner (i.e. *what* their task was), but rather whether it was their partner's turn or not (i.e. *when* they had to respond). This suggests yet another alternative, which is that individuals represent how the task is divided between self and other ("I respond when I see A or B, you respond when you see C or D"), but not what the specifics of the task rules are for their partner ("When you see C you respond with action 1, when you see D you respond with action 2). The discussion in this field tends to assume that only one of these possible representation contents is correct, so that co-representation always involves representing a certain type of information (e.g. stimulus-response mapping). However, we suggest that an alternative may be that the type of representation depends on the context. This discussion was originally raised in the adult literature, but it applies equally to findings on task co-representation in children. Indeed, the possibility of multiple types of representation content, requiring different levels of cognitive maturity, may be even more relevant for the developmental literature.

In fact, a similar question has arisen in the developmental joint attention research, regarding whether children represent simply *what* object has been shared in attention with another person, or also *how* it has been shared. Moll et al. (2008) showed that 14-month-old infants can correctly identify which of several objects an adult is attending to ambiguously based on exactly how they had shared attention to each of the objects just previously (i.e., based on where and how repeatedly they had found the objects, and the emotions, such as surprise, expressed with each one). Although this also raises the question of what the content of the

representation is, it differs from the joint action co-representation literature in that it does not assume that only one is correct. Instead, it assumes that different types of information can be represented depending on the context, or even the age of the individual. This idea has been taken from the huge body of work on perspective-taking, which demonstrates a dissociation between understanding *what* an observer sees (Level-1 perspective), versus *how* they see it (Level 2 perspective; Flavell, 1978). Much converging data now suggests that these levels of understanding show different developmental trajectories (Surtees & Apperly, 2012) and are subject to different signature limits (Surtees et al., 2012), supporting the theory that they involve distinct systems (Apperly & Butterfill, 2009).

The progress that has been made in perspective-taking and its influence on joint attention research could also be useful in the field of joint action. It may be profitable to consider the possibility that there is not just one type of information that is represented during a joint task, but rather that the type of representation differs depending on the context. Further, it may even be that there are different systems dedicated to different types of information. For example, it is possible that certain tasks may elicit co-representation of a partner's precise stimulus-response mappings, but others only elicit co-representation of information regarding which agent is responsible for which task set. If this is the case, comparing co-representation across different tasks and across development, as has been done in the perspective-taking literature, would be enlightening.

Top-down versus bottom-up processing. The actions of an individual can either be intentionally directed towards an object (top-down processing; Gregory, 1970) or reflexively drawn to an object as a result of an attention-getter (bottom-up processing, Gibson, 1966). This is not just the case for individuals, but also for interpersonal interactions involving objects.

For joint action, Knoblich, Butterfill and Sebanz (2011) argue that coordination may result from explicit and intentional efforts on the part of the actors, with a joint aim of achieving a joint goal (*planned coordination*). Alternatively, coordination may emerge spontaneously as a reflexive response to the actions of another person (*emergent coordination*). Emergent coordination could thus be seen as corresponding with bottom-up joint attention, in that it is elicited externally, whereas planned coordination is similar to top-down joint attention, which is initiated internally by an individual or group.

The processes associated with these two types of emergent versus planned coordination are likely to be different. For planned coordination, Knoblich et al. (2011) specify (at least) two processes: shared task representations and joint perceptions. These involve representing the task requirements of a partner, and representing their perceptions (e.g. their visual perspective), respectively. In contrast, for emergent coordination, they specify four possible processes: entrainment, common affordances, perception-action matching and action simulation. Entrainment is a social motor process, also described above, that results in individuals unconsciously synchronising with one another when making rhythmic movements such as walking together (Van Ulzen et al., 2008). Common affordances (Gibson, 1977) refers to the properties of an object that result in similar actions from two individuals, such as both producing similar grasping motions due to the shape and size of the objects to be grasped. Perception-action matching (Hommel, Muesseler, Aschersleben & Prinz, 2001) is another process that can lead to mimicry of another person, due to perception of another's actions resulting in activation of similar motor representations and a subsequent tendency to produce the same action. Lastly, action simulation is a process by which individuals form predictions about another's actions based on internal simulation of the action themselves (Gallese & Goldman, 1998). Knoblich et

al. (2011) argue that emergent coordination plays an important role in facilitating joint action, so discriminating between different processes that may be triggered by different contexts provides a more detailed picture of joint action as a whole as opposed to merely focussing on the initiation of top-down, planned actions. Furthermore, they also underline the potential for interaction between emergent and planned coordination for efficient joint action.

A similar distinction has been drawn for joint attention. Top-down processing may occur as a result of one individual attracting the attention of another to an object (e.g. via gaze leading: Bayliss, Murphy, Naughtin, Kirtikos, Schilbach & Becker, 2013; or communication: Carpenter & Liebal, 2011). Bottom-up processing may occur when an attention-getting event, such as an object falling, attracts the attention of both individuals and subsequently leads to sharing of attention to the event. Depending on which of these two scenarios individuals find themselves in, different steps of communication may be required in order to achieve joint attention (Carpenter & Liebal, 2011). In salient bottom-up scenarios, a single "sharing look" is likely sufficient to create joint attention, whereas in top-down scenarios one actor may need to perform several communicative steps in order to initiate attention with the other, direct her towards the object of interest and then, finally, to acknowledge the experience with a "sharing look."

Both the potential for multiple types of bottom-up and top-down processes and the possibility of interaction between these two forms of processing (as suggested by Knoblich et al., 2011, for joint action) could also be relevant for joint attention research. For example, it may be useful to investigate how joint attention is maintained or how it changes over the course of an interaction. For instance, an interaction could be initiated by bottom-up attention, but once this is established, top-down processes may play a role in maintaining the interaction. For example, in a bottom-up context, Edwards et al. (2015) found that adults whose gaze towards an object

had been followed by another person subsequently looked towards the other's eyes, thereby identifying that their gaze had been followed. Alternatively, attention could be initiated by top-down attention, but then use emergent processes to help to maintain the interaction. This may have the benefit, as it seems to in joint action, that less cognitive effort is required in order to uphold jointness and coordination across time.

Conclusions

We have highlighted how similar the two independent fields of joint action and joint attention are, despite their complexity – a complexity which is compounded by the fact that the divisions between them exist in both psychology and philosophy separately. All four areas have struggled with similar issues regarding how to define jointness, how to identify it using empirical work and how it is achieved by individuals involved in an interaction. Beyond the tricky issue of jointness, there are also other issues that bear significant resemblance across fields, including how to approach possible differences in the content of co-representations and the possibility of multiple systems, as well as distinguishing between top-down versus bottom-up processing in a joint context. In addressing each of these issues, we have highlighted ways in which bridging the gap between fields could be beneficial. One way of doing so is to adapt theoretical approaches that have been used in one field to see how they could apply to another, such as considering whether multiple types of representational content or different systems could be involved in joint action co-representation. Another way would be to share methods, such as implementing self-other interference paradigms in joint attention settings or identifying signature limits of joint action co-representation. Systematic comparisons such as we have presented here can help to promote the sharing of ideas across fields and also to identify ways in which different aspects of social coordination can interact with one another. This type of cross-talk between fields should

be encouraged more generally in cognitive science, and particularly between fields that clearly

have so many parallels to be drawn.

References

Bakeman, R. & Adamson, L. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interactions. *Child Development, 55,* 1278-1289.

Bayliss, A. P., Paul, M. A., Cannon, P. R. & Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review, 13,* 1061-1066.

Bayliss, A. P. & Tipper, S. P. (2006). Predictive gaze cues and personality judgments. Should eye trust you? *Psychological Science, 17*, 514-520.

Böckler, A., Knoblich, G. & Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 1404-15.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*, 2359-2367.

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review, 101*, 327-341.

Butterworth, G. (1995). Origins of mind in perception and action. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 29-40). Hillsdale, NJ: Lawrence Erlbaum.

Carpenter, M. & Call, J. (2013). How joint is the joint attention of apes and human infants? In J. Metcalfe & H. S. Terrace (Eds.), *Agency and joint attention* (pp. 49-61). New York: Oxford University Press.

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in

infancy. In A. Seemann (Ed.), *Joint attention: New developments in psychology,*

philosophy of mind, and social neuroscience (pp. 159-181). Cambridge, MA: MIT Press.

Cirelli, L. K., Einarson, K. M. & Trainor, L. J. (2014). Interpersonal synchrony increases

prosocial behavior in infants. *Developmental Science, 17*, 1003-1011.

Flavell, J. H. (1978). The development of knowledge about visual perception. In C. B. Keasey

(Ed.), *The Nebraska symposium on motivation: Vol. 25. Social cognitive development*

(pp. 43-76). Lincoln: University of Nebraska Press.

Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading.

Trends in Cognitive Sciences, *2*, 493-501.

Ganea, P.A., & Saylor, M.M. (2007). Infants' use of shared linguistic information to clarify

ambiguous requests for objects. *Child Development*, *78*, 493-502.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving,*

acting, and knowing: Toward an ecological psychology (pp. 67-82). Hillsdale, NJ:

Erlbaum.

Gräfenhain, M., Carpenter, M., & Tomasello, M. (2013). Three-year-olds' understanding of the

consequences of joint commitments. *PLoS One*, *8*, e73039.

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing

musicians. *Music Perception: An Interdisciplinary Journal*, *26*, 427-438.

Gregory, R. (1970). *The intelligent eye*. London: Weidenfeld and Nicolson.

Hamann, K, Warneken, F, & Tomasello, M. (2012). Children's developing commitments to joint

goals. *Child Development, 83*, 137-145.

Hamann, K., Warneken, F., Greenberg, J. R., & Tomasello, M. (2011). Collaboration encourages equal sharing in children but not in chimpanzees. *Nature*, *476*, 328-331.

Hommel, B., Muesseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC). *Behavioral and Brain Sciences*, *24*, 849-937.

Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, *13*, 522-525.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. *Psychology of Learning and Motivation-Advances in Research and Theory*, *54*, 59.

Leavens, D. A. & Racine, T. P. (2009). Joint attention in apes and humans. Are humans unique? *Journal of Consciousness Studies, 16*, 240-267.

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, *12*, 264-271.

Liebal, K., Carpenter, M., & Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. *Infancy, 15,* 545-556.

Hove, M. J. & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition, 27*, 949-960.

Milward, S., Kita, S. & Apperly, I. A. (2014). The development of co-representation effects in a joint task: Do children represent a co-actor? *Cognition*, *132*, 269-279.

Milward, S. J., Kita, S., & Apperly, I. A. (2017). Individual differences in children's corepresentation of self and other in joint action. *Child Development*, *88*, 964-978.

Milward, S. J. & Sebanz, N. (2016). Mechanisms and development of self-other distinction in dyads and groups. *Philosophical Transactions of the Royal Society B, 371*, 1-4.

Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science*, *10*, 826-835.

Moll, H., Richter, N., Carpenter, M., & Tomasello, M. (2008). Fourteen-month-olds know what 'we' have shared in a special way. *Infancy*, *13*, 90-101.

Neda, Z., Ravasz, E., Brechte, Y., Vicsek, T., & Barabasi, A.-L. (2000). The sound of many hands clapping. *Nature*, *403*, 849-850.

Piaget, J., & Inhelder, B. (1956). *The child's concept of space*. Routledge & Paul.

Plötner, M., Over, H., Carpenter, M., & Tomasello, M. (2015). The effects of collaboration and minimal-group membership on children's prosocial behavior, liking, affiliation, and trust. *Journal of Experimental Child Psychology*, *139*, 161-173.

Reddy, V. (2003). On being the object of attention: implications for self–other consciousness. *Trends in Cognitive Sciences*, *7*, 397-402.

Reid, V. M. & Striano, T. (2005). Adult gaze influences infant attention and object processing: implications for cognitive neuroscience. *European Journal of Neuroscience, 21*, 1763–1766.

Reid, V. M., Striano, T., Kaufman, J. & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *Cognitive Neuroscience and Neuropsychology, 15,* 2553-2555.

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*, 867-891.

Saby, J. N., Bouquet, C. A., & Marshall, P. J. (2014). Young children co-represent a partner's

   task: Evidence for a joint Simon effect in five-year-olds. *Cognitive Development*, *32*, 38-

   45.

Samson, D., Apperly, I.A., Braithwaite, J. & Andrews, B. (2010). Seeing it their way: Evidence

   for rapid and involuntary computation of what other people see. *Journal of Experimental

   Psychology: Human Perception and Performance*, *36,* 1255-1266.

Saylor, M.M., & Ganea, P. (2007). Infants interpret ambiguous requests for absent objects.

   Developmental Psychology, *43*, 696-704.

Searle, J.R. (1990). Collective intentions and actions. In P. R. Cohen et al. (Eds.), *Intentions in

   Communication* (pp. 401-415). MIT Press.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving

   together. *Trends in Cognitive Science*, *10*, 70-76.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own?

   Cognition, *88*, B11-21.

Siposova, B., Tomasello, M., & Carpenter, M. (submitted). Communicative eye contact signals a

   commitment to cooperate in young children.

Steinbeis, N. (2016). The role of self–other distinction in understanding others' mental and

   emotional states: Neurocognitive mechanisms in children and adults. *Philosophical

   Transactions of the Royal Society B*, *371*, 20150074.

Surtees, A. D., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of level-2

   perspective-taking in children and adults. *British Journal of Developmental

   Psychology*, *30*, 75-86.

Surtees, A. D., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, *83*, 452-460.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103-130). Hillsdale, NJ: Lawrence Erlbaum.

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioural and Brain Sciences, 28*, 675-735.

Tunçgenç, B., Cohen, E. & Fawcett, C. (2015). Rock with me: The role of action synchrony in infants' social and non-social choices. *Child Development, 86*, 976-984.

van Ulzen, N. R., Lamoth, C. J., Daffertshofer, A., Semin, G. R., & Beek, P. J. (2008). Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neuroscience Letters*, *432*, 88-93.

Venezia, M., Messinger, D. S., Thorp, D., & Mundy, P. (2004). The development of anticipatory smiling. *Infancy, 6*, 397-406.

Vesper, C., Butterfill, S., Knoblich, G. & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*, 998-1003.

Wenke, D., Atmaca, S., Höllander, A., Liepelt, R., Baess, P. & Prinz, W. (2011). What is shared in joint action? Issues of co-representation, response conflict, and agent identification. *Review of Philosophy and Psychology*, *2*, 147-172.

Wimmer, H. H. & Perner, J. J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children. *Cognition*, *13*, 103-103.

Wyman, E., Rakoczy, H., & Tomasello, M. (2013). Non-verbal communication enables

children's coordination in a "Stag Hunt" game. *European Journal of Developmental

Psychology*, *10*, 597-610.