# Explaining joint attention: Between epistemic justification and psychological processing

Lucas Battich[1,2]

[1]Institut Jean-Nicod, EHESS, CNRS, École Normale Supérieure, PSL Univesity, Paris, France
[2]Laboratoire de Neurosciences Cognitives et Computationnelles (LNC2), INSERM U960, École Normale Supérieure, PSL University, Paris, France

lucas.battich@ens.psl.eu

## Abstract

The ability to engage in joint attention, where two individuals attend to the same object or event together, provides an evidential basis for coordinated behaviours and interactions. To play this role, joint attention is often defined as a mutually open, or transparent relation between co-attenders. But how should this openness be characterised? Two broad theoretical views have been proposed. One view reductively accounts for the openness of joint attention in terms of individual mental states and properties. In contrast, according to non-reductive views, openness is based on some primitive intersubjective relation, irreducible to the individual states of each co-attender. I argue that tensions in these approaches arise from the methodological attempt to address normative and cognitive explananda simultaneously. Both approaches are primarily designed to tackle the normative epistemological concerns of joint attention, and their explanatory limitations arise when they extend their scope to psychological concerns. Instead, I propose adopting a cognitive-first methodological strategy. I outline the case for a probabilistic account of joint attention, and then assess its epistemic implications. The upshot is that the emphasis on a normatively justified state of joint attention may not be necessary for a psychological understanding of the phenomenon and its functional role.

**Keywords:** Joint attention; common knowledge; mutual awareness; social cognition

# 1. Introduction

Two people sit at a table with a piece of cake between them. They look at the cake, exchange glances, and smile. They are thus both attending to the cake. Importantly, they are both at the same time aware of each other's attention. If one of them were to say "Grandma made it", the referent would be clear to them both. This is a paradigmatic case of joint attention.

It is widely agreed that joint attention, in which two individuals attend to the same object or event together, plays an important role in language development and communication, in joint action, and in the progressive understanding that others can have different perspectives than our own. More generally, joint attention supports the development of mentalising, the ability to comprehend other people's mental lives (Carpenter et al., 1998; Moore & Dunham, 1995; Mundy, 2018). It is commonly assumed that a primary functional role for joint attention is to provide a basis for further coordinated activities between agents, from joint actions to shared beliefs and communication. Joint attention has thus been considered a necessary but not sufficient condition for individuals to engage in collective activities (Brinck & Gärdenfors, 2003), and plays an essential role in the constitution of the common ground involved in communication (Bruner, 1995; Clark, 1992). However, there is scant agreement on how joint attention should be analysed and how to explain its functional role. How does joint attention enable further coordinated behaviours and interactions between people? This paper aims to contribute to theoretical research on joint attention by examining how different methodological strategies have important implications for understanding what joint attention is and how its functional role is supported.

Most researchers agree that a key feature of the state of joint attention is that it is public or "out in the open" among the co-attenders. It is fully and immediately "transparent", or "mutually manifest", to them that they are jointly attending to the same object or state of affairs (Bakeman & Adamson, 1984; Calabi, 2008; Campbell, 2011; Carpenter & Liebal, 2011; Eilan, 2005; Peacocke, 2005; Tomasello, 1995). This feature puts the jointness in joint attention and distinguishes it from parallel attention, where we attend to the same object unilaterally, completely unaware of each other. John Campbell succinctly associates the openness of joint attention with this functional role:

> [W]hatever else is true of it, joint attention has an "openness" about it — there's some sense in which the situation is "open" to both attendees in a case of joint attention — in virtue of which joint attention ordinarily plays a distinctive role in rational, coordinated action. (Campbell, 2011, p. 417)

This notion of openness is largely metaphorical and indefinite. The challenge is to go beyond the metaphor. How should this notion be analysed? Following a distinction articulated by Campbell and others (Campbell, 2005, p. 288; c.f. León, 2021; Seemann, 2011; Watzl, 2012), most proposals fall into two broad theoretical views. One view reductively accounts for the mutual awareness characteristic of joint attention in terms of individual mental states and properties. In contrast, non-reductive views hold that mutual awareness is based on some primitive intersubjective relation, which is irreducible to the individual states of each co-attender.

I have two interrelated goals in this paper. First, I argue that part of the debate between reductive and non-reductive views arises through the conflation of two distinct explanatory aims: the normative aim of explaining under which conditions people are *justified* to be in a state of joint attention, and the cognitive aim of explicating the *psychological* capacities that are

involved in joint attention. I take a representative version of each reductive and non-reductive approach, and show that they are primarily concerned with the normative aim. Possible problems and explanatory limitations arise when they extend their scope to tackle the cognitive aim (section 2).

Second, and in contrast to their "normative-first" approach, I outline the case for a "cognitive-first" account of joint attention based on a probabilistic approach to mutual awareness (section 3). I suggest that mutual awareness in joint attention is not something co-attenders must *arrive* at, but that it is often implicitly assumed, by being responsive to a minimal set of cues, and that we must (un)learn that other people may not attend to the same things we attend. I conclude by returning to the normative concerns. I suggest that, in order to address the functional role of joint attention in enabling further social coordination, the notion of a well-founded and fully justified state of joint attention may be explanatorily expendable.

## 2. Methodologies in the reductive and non-reductive views

### 2.1. Explanatory aims

A primary functional role of joint attention is to provide a basis for further coordinated activities and interaction. The notion of a shared openness between co-attenders has been introduced to explain not only what joint attention *is* but also to address the question of its functional role. As Campbell (2011) notes, it is in virtue of such openness that joint attention rationalises coordinated activities and interactions.

To play this role, it is assumed that the state of joint attention itself must be well-founded or justified. There must be some strict epistemic conditions that a state of joint attention between two people should satisfy. Otherwise, it would not be able to play its functional role in supporting further social interactions. Even for philosophers who don't explicitly emphasise the role of joint attention for joint action and shared beliefs (e.g., Peacocke, 2005; Eilan, 2015), it is usually assumed that a satisfactory explanation of the openness of joint attention will involve stipulating the (if possible, necessary and sufficient) conditions under which two people are fully justified to be in such a state of openness: what makes a person having immediate justification for taking their and another's attention to be of the same object?

On the other hand, joint attention is a psychological phenomenon, and any theoretical analysis will be constrained by psychological plausibility. For example, Christopher Mole (2017) notes that, "since joint attention is achieved by young children, its achievement cannot plausibly be thought to make any sophisticated intellectual demands" (cf. Roessler, 2005). We need to spell out what the mental and cognitive requirements for joint attention are: how is openness achieved? An account of the openness of joint attention is therefore seen as a necessary step in addressing these two different explananda:

(i).   Normative: What are the epistemic conditions that must hold for a phenomenon to be a well-founded and justified instance of joint attention?
(ii).  Cognitive: What cognitive capacities and mental processes or understanding are involved in joint attention, which even infants and young children may be capable of having?

These two questions are often tackled at once in the current debate. Supposedly, the openness of joint attention must be explained in such a way that it is epistemically well-founded and justified, and simultaneously sheds light on the mental wherewithal necessary to achieve said

openness. Answering the epistemic question may require some minimal talk about the mental processes involved in joint attention. Ideally, this epistemic-first strategy would try to identify some of the processes and capacities necessary to achieve a well-funded and fully justified state of joint attention.

Separating explananda in joint attention is not entirely new. Naomi Eilan has previously noted that

> there is the question of the kind of mutual awareness we find in joint attention, in virtue of which we think of minds as being transparent to each other (call this the 'Epistemological Question'). And there is the question of the kind of understanding of attention which a child at the age of 1–2 might plausibly be thought of as possessing and bringing to bear on these situations (call this the 'Concept Question'). (Eilan, 2005, p. 4)

Eilan's distinction doesn't exactly overlap with the one above, as the cognitive question is not restricted to conceptual understanding of attention, but rather encompasses possible non-conceptual, non-propositional understanding, and subpersonal psychological processes. Nevertheless, whether we take a normative-first or a cognitive-first approach when analysing joint attention has important implications. I will examine the contrast between reductive and non-reductive accounts of joint attention in light of these two different explanatory aims. Both accounts take the normative question (i) as a starting point and motivation, and attempt to address the cognitive question (ii) from there. In the rest of this section, I illustrate possible issues that arise from their epistemic-first methodology. For a more tractable treatment, I focus on particular versions of the reductive and non-reductive approaches to joint attention: the iterative account and the relational account, respectively. My intention is not to provide any comprehensive criticism of these accounts, as they have been offered elsewhere (see León 2021; Battich and Geurts 2021), but rather to highlight the explanatory limits inherent to the epistemic-first methodology strategy.

## 2.2. The reductive approach: iterative account

According to the reductive approach, the openness in joint attention can be explained in terms of the individual mental states of each co-attender. While this view does not deny the experiential or phenomenal character of joint attention, it holds instead that this character is better accounted for in terms of individual epistemic states. Reductive views may come in principle in many forms. Historically, however, the prominent reductive views tend to analyse the openness of joint attention in terms of common knowledge or similar notions such as mutual awareness, constructed in line with the analyses from Lewis (1969) and Schiffer (1972). On this definition, A and B are jointly attending to $x$ when they are mutually aware, or enjoy common knowledge, that each of them is attending to $x$. Common knowledge is defined as giving rise to iterative structures like the following:

$p$ is common knowledge between A and B iff

- A knows that $p$,
- B knows that $p$,
- A knows that B knows that $p$,
- B knows that A knows that $p$,
- and so on, *ad infinitum*.

On this iterative construal, joint attention involves nested psychological states that both A and B would have to entertain about each other. This is the hallmark of recursive mindreading, where one subject (A) attributes mental states to another (B knows that …), which, in turn, refer to the first subject's own mental states (B knows that A knows that …). Most criticisms of this approach centre on the requirement of recursive mindreading (León, 2021). One argument from phenomenology notes that the openness in joint attention is immediate and effortless, so it is implausible that joint attention requires recursive mindreading (e.g. Gallagher, 2011). Related arguments point to the computational complexity of recursive mindreading, arguing that it is cognitively demanding and psychologically implausible even for adults (Campbell, 2018; Eilan, 2005).

Felipe León (2021) has accurately summarised the problems of recursive mindreading and the iterative approach to joint attention[1]. My point here is that the claim that reductive accounts necessitate recursive mindreading already conflates normative and cognitive explananda. To tease out how these differ, I focus on a key argument against the reductive approach, which appeals to "coordinated-attack" scenarios (Campbell, 2005, 2018; Wilby, 2010).[2]

> *Coordinated-attack scenario*: Two individuals in separate booths must both attack the same target among many, at the same time. For this, they will have to coordinate their individual actions. Additionally, there is always a non-zero chance of distorted communication between booths, so that when one individual has chosen a target and communicates this to the other, they will not know for sure whether the communication has been received. Supposing that the second individual does receive the message specifying the target, they could, in turn, send a message back to confirm receipt of the target. But, again, they will not be sure whether this confirmation has been received.

Coordinated-attack scenarios are often assumed to show that no finite iteration of inferences will allow the participants to engage in rational coordinated attack and secure victory. Yet in normal situations, where we are both present in the same physical space, we can easily arrive at a successful coordinated outcome. When the two individuals and the target are all co-present, as in most normal circumstances, then "everything is out in the open to such an extent that we can rationally attack" (Campbell, 2005, p. 292).

For Wilby (2010), the coordinated-attack problem "highlights that once one gets embroiled in *supposing* that an act of transparent communication or shared knowledge requires a set of hierarchical to-ing and fro-ing about who knows what, then there will be no end to the matter"

---

[1] León offers four arguments against recursive mindreading in joint attention. The argument from phenomenology points out that joint attention often feels effortless and immediate, without any consciously performed inferences or iterations. The argument from cognitive demandingness highlights the cognitive difficulties in carrying out recursive inferences. The argument from perspective-taking notes that infants seem to engage in joint attention before acquiring the perspective-taking capacities required for recursive mindreading. The argument from normative force relies on the "puzzle of coordinated attack". As León assumes that all reductive views must necessarily endorse recursive mindreading, he takes these arguments to count against such views. It is this assumption that I would like to bring into question.

[2] The coordinated-attack problem was first introduced in its current form by Akkoyunlu et al. (1975) in the context of dynamical systems engineering and, while it soon became a key fixture in epistemic logic, it had little to do with mental representations and psychological processes.

(my emphasis). Going up only two or three levels up the recursive chain will not suffice. As Wilby (2010) and Campbell (2018) note, for any level in the chain, that level does not lead to openness or mutual awareness, or else will require a further step in the recursive chain[3]. The conclusion of the coordinated-attack argument is what Wilby calls a disastrous "paradox" for the reductive approach: mutual awareness requires an infinite recursion of overlapping mental states, but this requirement is psychologically implausible.

There are two points where the argument from coordinated-attack fails. First, there is nothing in the coordinated-attack scenario that implies that the iterations must be actually *represented* in the mind of each individual in a case of joint attention. Presupposing that each iteration must be represented already relies on the convergence of both normative and cognitive aims described above. Second, this rendering of the "paradox" assumes that, on the reductive approach, common knowledge is the *result* or *endpoint* of the iterations. On this assumption, the openness in normal joint attention scenarios would be equivalent to the openness that the individuals in our separate booths are infinitely pursuing. This would spell problems for the reductionist: if a chain of iterations does not produce common knowledge in the coordinated-attack scenario, it cannot explain how the openness of joint attention is achieved. Yet the reductionist is not committed to making such an assumption in the first place. The view that a reductive approach to joint attention involves recursive mindreading is based on normative analyses of common knowledge such as those by Lewis and Schiffer. Traditionally, their approach presumes that there are situations with some *finite* conditions, out of which the infinite iterations logically follow. What makes a situation a common knowledge situation are those finite conditions, not the recursive iterations themselves, nor their endpoint. In other words, the iterations arise as logical implications from some given finite situation, and are not necessarily represented in anyone's reasoning (Lewis, 1969, p. 53). Coordinated-attack scenarios only show that the individuals in their separate booths lack the appropriate finite conditions. But what are these conditions?

To date, providing a good account of those finite conditions has proved somewhat problematic, but only when, in addition, such conditions must also account for the psychological processes and mental wherewithal necessary for joint attention or common knowledge. If we do away with this cognitive explanandum, we are left with only the normative question. On this question, even when acknowledging that joint attention is a psychological *phenomenon,* the openness of joint attention can be seen as a purely normative epistemological notion, without any direct commitments to particular psychological *mechanisms*. Analyses of joint attention and mutual awareness in the style of Schiffer's and Lewis' are not necessarily committed to recursive mindreading or any regression of mental states. This conclusion leaves unanswered what mutual awareness is, psychologically speaking, as a mental state enjoyed by both adults and infants. In other words, the unfortunate upshot is that the cognitive explanandum of joint attention is left unaddressed.

One example of a purely normative treatment of phenomena that share the openness under consideration is Bart Geurts' (2019a, b) commitment-based conception of communication and

---

[3] One may suggest that people just reason two or three levels up the hierarchy and then conclude their knowledge is shared. This is an empirical, not a logical, suggestion; and one with no evidential support (e.g., Liddle & Nettle, 2006; Thomas et al., 2014). The approach implicit in this suggestion is normative-first: start with a logical, normative theory, and truncate it to fit what (we think) humans can do. I suggest adopting instead a cognitive-first approach and evaluate any normative implications thereafter.

common ground. Geurts uses the notion of shared commitments to explain these phenomena. When a commitment to act on *p* is shared and accepted between agents, it thereby becomes part of their common ground, allowing further action. But he stops short of postulating the psychological processes that must accompany the normative theory. On this framework, it is not necessary to psychologically represent our commitments. If we are committed to act on *p*, and *p* entails *q*, then we are also committed to act on *q*, even when we are unaware of the entailment. Building on Geurts' framework, Scarafone (2024) has recently developed a commitment-based account of joint attention, where "*a* and *b* jointly attend to *x* if and only if each of them shares a commitment with the other to attend to *x* and they behave accordingly" (p. 290). He explicitly notes that this is "a normative account, in both attitude and subject matter". This doesn't mean that one cannot propose psychological hypotheses based on such normative accounts. Scarafone suggests that the commitment-based account can allow for the formulation of scientific hypotheses on the different psychological resources and behavioural patterns involved in joint attention. This is a suitable methodological move, which could lead, for example, to a better empirical understanding of the sense of commitments that joint attenders may experience. But the important point is that it is possible to provide an iterative reductive account of joint attention, addressing the normative question (i), separately from any account of psychological processing and mechanisms directed at the cognitive question (ii). The iterative approach is purpose-built to address the normative question, and – even when acknowledging that joint attention is a psychological phenomenon – *by itself* does not inform us about which psychological processes and capacities are behind joint attention.

## 2.3. The non-reductive approach: relational account

Prominent non-reductive approaches tend to focus on the experiential aspects of joint attention (Campbell, 2005, 2018; Seemann, 2019; Wilby, 2010, 2023). Joint attention is normatively individuated from other mental and social states by virtue of its particular phenomenal character, and may thus play an epistemic role in justifying shared beliefs and joint activities. On this approach, there is a primitive intersubjective relation behind joint attention, in the sense that it cannot be analysed any further in terms of its individual components. John Campbell thus proposes that joint attention is a primitive type of conscious state: "just as the object you see can be a constituent of your experience, so too it can be a constituent of your experience that the other person is, with you, jointly attending to the object" (2005, p. 288, cf. 2018). Naomi Eilan argues that joint attention is grounded in experiences of "you-awareness" and "communication-as-connection", which are primitive conscious states (2015). Following Campbell, Axel Seemann (2019) argues that our perceptual experience during joint attention is a primitive joint state. What each of us experiences during joint attention cannot be reduced to our individual psychological states, but is determined by the spatial organisation between us and the common object of our attention (Seemann, 2019, p. 75).

Given its prominence in the philosophy of joint attention, I'll focus here on Campbell's relational analysis. This view is based on Campbell's relational theory of perception: perception consists of a non-representational relation between the perceiver and the object or scene being perceived (Campbell, 2002).

> On a relational view, joint attention is a primitive phenomenon of consciousness. Just as the object you see can be a constituent of your experience, so too it can be a constituent of your experience that the other person is, with you, jointly attending to the object. (Campbell, 2005, p. 288)

Campbell's analysis is based on the premise that joint attention can be explained fully in terms of perceptual experience, and thus is not susceptible to an explanation in terms of the knowledge, beliefs, or awareness of the two participants (Campbell, 2018, p. 120). For this reason, it has proved to be an attractive theoretical position for cognitive and developmental psychologists (e.g. Moll & Meltzoff, 2011; Hobson & Hobson, 2011). One serious criticism of this approach is that it simply embeds in the analysis of openness or mutual awareness the property that is to be explained, i.e. the openness of joint attention (Peacocke, 2005, p. 300). Further problems can arise due to its reliance on a relational view of perception and the narrow focus on perceptual experience. Every account admits that you could mistakenly take yourself to be in joint attention when in fact you are not. But, since the non-reductive approach allows for dissociations between one's perceptual experience and one's beliefs about that experience, it seems to allow for the converse, puzzling possibility: that you take yourself to be attending alone to *x*, when in fact you are jointly attending to it with someone. Here, you *are* engaged in joint attention and having the relational experience of openness, yet you are not aware that you are so engaged or so experiencing. But this conclusion seems to go against the key feature of joint attention that the view attempts to explain: co-attenders are mutually aware of their joint attention[4] (see Battich & Geurts, 2021).

As is perhaps already evident, the key motivation for non-reductive approaches is to address the normative explananda of joint attention. Campbell, as well as others who have expanded on his relational theory, are explicit that their aim is to explain the normative role of joint attention in rationally justifying coordinated actions and other joint activities (cf. Vincini, 2024; Seemann, 2024). The limitations of this view arise when it is also taken to provide insight into the cognitive question, as the openness of joint attention becomes merely a description of phenomenal aspects of experience. Joint attention is justified between two co-attenders in virtue of them being in the appropriate phenomenal state. Being primitive, this state cannot be further explained. But neither does it inform us about the psychological capacities behind joint attention. The unfortunate upshot, again, is that the cognitive explanandum of joint attention is left unaddressed.

Again, it is possible to use the relational account of joint attention as a framework to propose psychological hypotheses. Endorsing a relationalist approach, Vincini (2024) suggests that some psychological processes involved during joint attention include "commonplace psychological processes of pattern formation", associative processes, predictive processes that categorise the sensory input that one encounters, or through acquired experiential patterns, and perhaps even statistical learning (Vincini, 2024). These processes are said to underlie the

---

[4] A disjunctivist approach to experiences, such as Campbell's, could accept this possibility and argue that it does not hinder the epistemic justification conferred by joint or solo attentional experiences. Disjunctivism holds that the experiencing subject is necessarily not in a position to distinguish experiences of different kinds (veridical perception, hallucination, etc.) simply by virtue of having the experiences (McDowell, 1982; Haddock and Macpherson, 2008). Similarly, Social disjunctivism holds that, when subjects have an experience of joint attention, they cannot distinguish a "good" experience (e.g., there is a co-constituent) from a "bad" experience (e.g., no co-constituent) (Seemann, 2019). Yet, this still doesn't undermine the idea that one's experience is epistemically justified in the "good" joint case (McDowell, 2008). Further, the social disjunctivist can allow that, when subjects have an experience of *solo* attention, they cannot distinguish a "good" solo experience from a "bad" experience (e.g., there *is* a co-constituent), and this still would not undermine the epistemic justification of one's experience in the "good" solo case.

typical set of behavioural attitudes and responses between infants and caregivers, such as gaze alternation, convergent bodily gestures, and pointing. However, such general psychological capacities as pattern formation, predictive processing, categorisation, and associative and statistical learning, could equally be said to underlie a reductive commitment-based theory of joint attention. This doesn't constitute per se a criticism of Vincini's psychological view. In fact, it is extremely likely that joint attention includes such general processes as pattern recognition, prediction, categorisation, and learning (see section 3.1). The point I want to highlight is that this description of psychological processes does not follow in a principled or straightforward manner from the relational view *itself*.[5]

A description of a primitive intersubjective relation between co-attenders and its phenomenology may well provide a satisfactory answer to the normative question: what conditions hold for a phenomenon to be a well-founded and justified instance of joint attention. In particular, it addresses the problem of how joint attention serves as a categorical and rational basis for joint actions, which forms the explicit motivation of several non-reductive approaches (Campbell, 2005; Wilby, 2023; Seemann, 2024). The important point is that this description is, itself, as yet silent on the subpersonal psychological states and processing of each individual (cf. Schmitz, 2014, p. 239). As with the iterative account, the relational approach is purpose-built to address the normative question, and – even when acknowledging that joint attention is a psychological phenomenon – *by itself* does not inform us about which psychological processes and capacities are behind joint attention.

This conclusion is further supported by the conceptual equivalence between iterative and primitivist analyses. For primitivists like Campbell (2005) and Seemann (2019), the state of mutual awareness in joint attention is a factive state. You cannot be aware that the other person is currently co-attending with you to the same perceptual target, unless that person is, in fact, a co-attender. The factive character of the mental states of each co-attender is also usually assumed for the reductive, iterative approach. Two people are mutually aware of $p$ only when both of them are equally justified in following the infinite logical implications of their joint epistemic state. Interestingly, Wilby (2010) has shown that the factivity assumption implies that the iterative and relationalist views are conceptually equivalent — at least under some versions of each view. In particular, the equivalence holds for Schiffer's analysis of common knowledge, commonly taken as a paradigm of the reductive, iterative approach when applied to the openness in joint attention. Wilby (2010) proved that Schiffer's analysis of common knowledge turns out to be logically equivalent to an analysis including a single intersubjective element irreducible to the mental states and properties of each individual, that is, a relational state irreducible to the individual cognitive states of the individuals in question (Wilby, 2010, p. 92). It is important to be clear about the implications of such equivalence. The equivalence holds because both analyses are restricted to the normative aim: how is a fully justified and well-founded accurate state of joint attention attained? But the logical equivalence between the two accounts doesn't necessarily shed any light on the cognitive aim: what are the psychological processes and capacities underlying joint attention? Indeed, their unreconcilable differences seem to arise only when they are taken to address, post factum, the cognitive question regarding the psychological processes behind joint attention.

---

[5] It is also salutary that Vincini (2024), along with Geurts (2019a, b) and Scarafone (2024), are explicit about the aims and explananda of their respective accounts.

To recap, in this section I suggest that both iterative and relationalist accounts are purpose-built to address the normative question primarily. While they acknowledge that joint attention is a psychological phenomenon, the iterative and relationalist accounts do not have, by themselves, direct commitments to any particular view on — and don't directly inform us about — which psychological capacities and processes may sustain joint attention. The interim point is that, on these accounts, the cognitive explanandum of joint attention is left unaddressed or underdeveloped. If we are interested in the psychological states of real humans and children, then we should at least acknowledge the explanatory limitations of such "normative-first" approaches to joint attention.

## 3. A cognitive-first approach to joint attention

### 3.1 Joint attention as a probabilistic psychological problem

In 2003, Michael Tomasello remarked that child language acquisition is not a logical problem, but a psychological one. He urged that a theory of human linguistic competence should be based less on analogies to formal languages and more on research in the cognitive sciences (2003, p. 328). Unfortunately, Tomasello himself didn't fully apply this dictum to the topic of common knowledge and joint attention, assuming, along with many others, that it must involve either something akin to recursive mindreading, or some primitive non-analyzable plural "we" subject (cf. Zawidzki, 2013):

> From early on as well, infants communicate with others referentially, inviting
> them to jointly attend to something, and this requires recursive inferences
> about mental states embedded in mental states. (Tomasello, 2019, p. 44)

Given the "normative-first" focus of traditional approaches to joint attention, however, I propose that, if we are interested in the cognitive question, there are no strong reasons to assume a priori that the key feature of joint attention is a fully symmetric epistemic or phenomenal openness between co-attenders. Instead, I propose that a more fruitful approach to describe the triadic interaction of joint attention is to concentrate on what factors or aggregate of factors each individual co-attender is responding to, so that this interaction can be established, without presuming in advance the nature of the epistemic or phenomenal sophistication they must achieve. In this section, I provide the outline of a "cognitive-first" approach to assess the jointness of joint attention, where there is no emphasis on any *justification* for the mental states of an individual during joint attention.

My proposal here is limited to a general sketch of the computational steps behind join attention, compatible with different proposals regarding the mechanisms that support this computation. In this and the following subsections, I pursue this sketch in a two-tier approach and identify some of the capacities that a cognitive agent would require in order to perform such computations. In specifying a computational sketch at a high level of explanation, I only intend to use computation as an explanatory and descriptive strategy, remaining as neutral as possible regarding realist claims about whether the mind is or not a computing system. Specifically, I am not committed to any strict reduction between this computation and the mechanisms in the brain that may support it, so that theories that propose non-computational and even non-representational mechanisms at lower levels of explanation should be equally well accommodated in this picture. Finally, in section 3.3, I will use the framework developed to see how we can bring back the normative concerns, and discuss two possible worries.

To start, engaging in joint attention requires one to know what one is attending to, as well as what the other is attending to. This in turn requires the combined processing of different types of information: information about one's own attentional state (including interoceptive and proprioceptive information; Mundy & Jarrold, 2010); information about the other's attentional state; and information about the target of joint attention (Mundy, 2018; Siposova & Carpenter, 2019; Battich et al., 2020). All three types of information come under degrees of uncertainty and have to be combined in order to establish whether we are attending to the same object together. Extending Siposova and Carpenter's approach (2019), the computational challenge in establishing joint attention can then be specified as involving the estimation of three kinds of probabilities:

(P1). Degrees of certainty about the object of my perceptual attention.
(P2). Degrees of certainty that the other attends to the same object as me.
(P3). Degrees of certainty that the other's attention is being shared with me.

We can assume that the first kind of certainty is high in most situations, at least the ones lending themselves to joint attention; though that is of course not necessarily the case. On the proposed hypothesis, estimating probabilities (P2) and (P3) can be more cognitively demanding — though such processes can of course (and likely do) occur unconsciously and pre-reflectively, at the subpersonal level. To estimate these probabilities, I need ways of gathering information about the other's attentional state and the extent to which they share that state with me. As Siposova and Carpenter (2019) note, different ways of gathering information require different levels of assumptions. For example, if I hear, but not see, that my friend is near me, I may assume they are physically present, and therefore that they do perceive object $x$ in front of us. But this is only an assumption, and the certainty in probabilities (P2) and (P3) can only be estimated up to some point.

By using the terminology of "probabilities", I here intend to refer to uncertainties in the cognitive system, where "the process of probability estimation" refers to processes of dealing with such uncertainties. There are several ways in which the system can deal with these uncertainties, and thus of estimating the probabilities (P2) and (P3). Such estimation likely involves mutually inclusive processes of learning, prediction, and (heuristic-based) assumptions. For example, during development I can *learn* that when I hear, but don't see, that you are nearby, I can readily *assume* that you are physically present. This assumption allows me to *predict* that you also see object $x$ in front of us, regardless of whether this prediction is correct or not. More precisely, I don't consciously do any of these things, but more likely subpersonal psychological systems engage in learning, assuming, and predicting.

Importantly, we don't need to suppose that in order to engage in joint attention agents must always attempt to arrive at a high or even perfect level of certainty regarding these estimates. This basic probabilistic element, where imperfect probabilistic estimates may be sufficient to engage in joint attention, constitutes the first tier of my proposal. In section 3.3, I suggest that this element may be enough to question the explanatory need for a fully justified and well-founded state of joint attention. Ideally, however, this simple probabilistic account will be characterised further. After all, estimating probabilities (P2) and (P3) could presumably involve a variety of processes and capacities. What follows is a basic suggestion about the possible steps involved, resting on recent psychological literature, and making the second tier of my proposal.

## 3.2 Mutual awareness is implicitly assumed

As just noted, it is not necessary that agents estimate these probabilities always with high certainty. A minimal heuristic-like step would be to implicitly assume, in a rough and ready manner, that we already *are* attending together to the same target. In a nutshell, I suggest that mutual awareness in joint attention is neither a primitive nor reductive intersubjective relation that co-attenders must arrive at, but that it is often implicitly assumed, and that we must (un)learn that other people may not attend to the same things we attend, or may not share the same perceptual knowledge we are currently enjoying. On this proposal, when we assume that we are sharing attention to the same target, we have a prior tendency to give a high value to both probabilities (P2) and (P3), and we must learn to properly estimate these probabilities in different contexts and situations.

On the view I propose, then, to jointly attend to an object is to pre-reflectively assume that one's attention to the same object is shared with someone else's. This implicit assumption is not necessarily rigidly automatic, but context-dependent. During development and through lived experience, we learn to be sensitive to the contexts and situations where we can default to the assumption of sharing attention to the same object, and those where we have to curb down this assumption. Contrary to the traditional reductive iterative view, attaining perceptual common knowledge towards the same object is not cognitively taxing, but curbing it down could be: taking into consideration whether other people do not share your object of attention could be cognitively demanding, at least during development and in novel situations with no prior precedents. Unlike non-reductive relational views, which posit some intersubjective primitive phenomenon behind joint attention, the view I propose is anchored on the individual. It concerns the mental processes that an individual A must go through in order to say that she is jointly attending with B to *x* (and takes herself to be mutually aware with B about so being in joint attention to *x*).

This account follows from the recent "altercentric hypothesis" of infant development (Southgate, 2013, 2020; Grosse Wiesmann & Southgate, 2021). This hypothesis holds that infants are inherently other-centred or altercentric: from very early in life have a tendency to perceive the world through the perspective of others, and are thus able to predict and interpret their behaviours (for a review of the empirical support for this hypothesis, see Grosse Wiesmann & Southgate, 2021; Southgate, 2020).

Traditionally, research in developmental psychology has followed a picture of infant development consisting of a slow progression of understanding different facets of perspective-taking and of predicting others' behaviours and mental states, culminating in the ability to understand others' false beliefs (Southgate, 2020; Wellman et al., 2001). This trajectory assumes that infants struggle to disregard their own perspective, which brings difficulties in a task that requires an understanding of what the other can see. This idea of perspectival conflict is meant to explain the results of the traditional false-belief task, where 3-year-olds fail to predict the other's actions based on understanding what the other can and cannot see (Wellman et al., 2001).

However, this traditional view has been increasingly empirically challenged. Infants as young as 3 months seem not only sensitive to what others can see and can track their attention (Scaife and Bruner, 1975) but, crucially, by 6 month of age they also use this information to interpret and predict other's actions (Choi et al., 2018; Luo and Johnson, 2009; Ting et al., 2021. Kim and Song, 2015). Southgate, Grosse Wiesmann and colleagues suggest that, in early social interactions, instead of using demanding processes of flexible perspectival-conflict resolution,

infants "*primarily* represent the other's perspective and, in effect, abandon their own perspective" (Grosse Wiesmann & Southgate, 2021):

> During infancy, it could be an adaptive strategy to direct limited attentional resources to the altercentric perspective rather than their own representation of events in the world (Southgate, 2020). Especially, in situations where the other's actions and communication are likely to be relevant, encoding the situation from the other's perspective would be an efficient way to predict and understand their behavior. (Grosse Wiesmann & Southgate, 2021)

The account of joint attention I propose here follows this developmental hypothesis: infants preferentially encode objects and events that others perceive, and generate expectations about others' behaviours assuming that their perception towards such objects or events is shared. In other words, the probabilities (P2) that the other attends to $x$, and (P3) that the other's attention to $x$ is being shared, are initially assumed to be high. During development, infants learn to better estimate these probabilities in different contexts and situations.

This proposal already goes some way into addressing the cognitive question: what cognitive capacities and mental processes or understanding are then involved in joint attention? We can go further: what capacities and mental processes are involved in estimating probabilities (P1) to (P3), even when making (possibly incorrect) assumptions about them? Addressing this question does require some normative ground-laying, although of a very different sort than that exemplified in what I called the normative question above. What minimal capacities are *necessary* to pre-reflectively assume that one is sharing attention to the same object with others (regardless of whether that assumption is justified or not)? We can start with a set of minimal cues that I as a co-attender should be capable of recognising:

(a).  You engage with the world as I do.
(b).  You are interacting with me.

A creature capable of understanding as much, should be capable of arriving at the assumption (even if erroneous) that they are sharing their perceptual engagement towards an object with someone else. This cognitive process could be paraphrased as "I have a certain relation to $x$, and since you and I are so similar, and you are interacting with me, (I assume) you have it too."

When I assume mutual awareness, this does not imply that this assumption must be fully conscious, reflective or deliberate, or that I have to consciously entertain the proposition that you are similar to me. It does not even require having a concept of mutual awareness. The only requirement is that I recognise and respond to the cues that you provide by implicitly assuming that you engage with the object in the same way I engage towards it. Such recognition and response are plausibly supported by subpersonal sensory-motor and affective processes (cf. Reddy, 2010; Gallagher, 2010). Interaction is a key component here. Without at least a minimal form of interaction (or the assumption, from my point of view, that you interact with me), it seems implausible to assume that I engage towards some object *together* with someone else.

The abilities to recognise (a)-(b) and the assumption of joint attention derived from them do not necessarily have to lead to conceptual and reflective awareness, in that both infants and

adults do not need to consciously reflect on them.[6] Of course, I could become reflectively aware of our joint engagement towards *x*. Usually, this occurs when the assumption of joint attention breaks or misfires. If I say to you "Grandma made it", and you show no comprehension of the intended referent, I can become retroactively conscious that I assumed, incorrectly, that we were looking at the cake together. When interaction fails, an individual may learn to tone down their assumptions of openness in similar future interactive situations, and reevaluate the set of cues that trigger their assumption in those situations. Put differently, through error they learn to estimate the probabilities (P2) and (P3) above more efficiently in those situations.

What is involved in the abilities to recognise (a) and (b)? One general agreement is that, in order to recognise that (a) you engage with the world (in a potentially similar manner to how I engage with it), I need to recognise you as a minimal intentional agent. Disagreements start to pile up on what this recognition consists of. Tyler Burge argues that to engage in joint attention, there is no need to represent others' psychological states; merely attributing purposive behaviour is sufficient (2011), an ability which might emerge as early as 6 months of age (Woodward, 1998). Vasudevi Reddy suggests that affective engagement is paramount for experiencing others as purposive agents, and proposes a developmental trajectory where triadic attention recognition (responding to others' attention to *things*) seamlessly follows from more basic forms of dyadic attention recognition (responding to others' attention to *self*) (Reddy, 2010, 2011; cf. Trevarthen & Aitken, 2001). Importantly, Reddy maintains that around 4 to 5 months of age, dyadic attention recognition already has the basic form of object-based joint attention, since the other's attention is understood as being directed at the infant's own body (Reddy, 2011). In contrast, according to Tomasello, infants begin to recognise others as intentional agents around 8 to 9 months of age when they differentiate means from ends in shared goal-directed activities (Tomasello, 1999). Tomasello insists that understanding the other as an intentional agent during joint attention is supported through recursive mindreading abilities, which marks a developmental shift from earlier dyadic interactions (Tomasello, 2018, 2019).

In a recent review of the developmental evidence, Henrike Moll and colleagues show a middle ground between Reddy's and Tomasello's accounts (Moll et al., 2021), which provides support for the proposal sketched here. They show that, contrary to Reddy's claim, there is no evidence that infants in early dyadic interactions contemplate the other's perception of themselves (and their bodies or actions) as an object of joint engagement. Contrary to Tomasello's claims, they review evidence suggesting that the transition from dyadic to triadic engagement does not occur as a sudden shift involving recursive mindreading abilities. Moll et al. (2021) suggest that the capacities to engage in joint attention, rather than being there all along or emerging suddenly, arise in a gradual fashion from dyadic to triadic relations. The minimal understanding of intersubjectivity already at play during dyadic exchanges as early as 2 months of age, still plays a key role during later triadic joint attention, as infants and their attention partners increasingly engage in "sharing looks", nodding, and other proto-communicative behaviours (Carpenter & Liebal, 2011; Hobson & Hobson, 2011). The recognition that (b) you are interacting with me may arise in a minimal form even earlier, as infants shortly after birth preferentially seek intersubjective interactions (Trevarthen, 2011). Hobson and Hobson suggest that

---

[6] In contrast to the analyses of Campbell (2005), Peacocke (2005), Wilby (2023), and others, I do not aim to characterise the phenomenal experience of openness itself. I remain noncommittal to what the phenomenology of this assumption might be. However, it is important to note that the functional role of assuming mutual awareness, in the sense described here, can be carried out without invoking its phenomenological aspect.

the same tendency of infants only weeks after birth to orient towards others, is what grounds the tendency to identify with the attitudes of other people during joint attention (Hobson & Hobson, 2011).

The conceptual account I propose here is compatible with Moll et al.'s (2021) reading of the developmental evidence (cf. Moll, 2024). The situations that prompt the assumption of joint attention may be at first crude and minimal, and refined throughout development. Infants display a variety of behaviours during the developmental trajectory between dyadic and triadic engagements, such as pointing gestures and gaze alternation, which show they are progressively sensitive to situations where joint attention is *not* yet attained, or where it is yet uncertain (Carpenter et al., 1998; Mundy, 2018).

The impulse to assume shared experiences or shared knowledge is not restricted only to infants during development. A similar phenomenon is observed in adults, where someone's own knowledge will affect, however implicitly, their ability to reason about another person's beliefs (Epley et al., 2004). Referred to as the "curse of knowledge", people are egocentrically biased to assume that others know what they themselves know (Birch & Bloom, 2007; Farrar & Ostojić, 2018). There is as yet no empirical parallel to these findings for joint attention. However, even though adults are proficient mind-readers, in order to engage in joint actions and behaviours based on joint attention, it doesn't seem necessary for them to always deploy their full cognitive efforts and achieve a fully justified and logically well-founded state of joint attention. On the contrary, adults seem to be sensitive to situations where they can allow themselves to make the (possibly incorrect but effortless) assumption of sharing attention with someone to *x*, and situations that require them to put more cognitive effort into their estimates of probabilities (P2) and (P3). The coordinated-attack scenario would be an extreme and unnatural example of the latter situations.

To recap, the interim point is that engaging in joint attention involves recognition of a minimal set of cues (a) and (b), and the biased (and possibly fallible) computations of probabilities (P2) and (P3). We may not even have to arrive at any high degree of certainty about these probabilities, but often implicitly assume, in a low-effort heuristic-like step, that we share engagement towards the same target. This assumption is of course not automatic, incorrigible, nor context-independent, but likely depends on the cues gathered in each context and situation, and is usually dynamically corrected and adjusted. It is proposed in contrast with the notion that we have to *achieve*, through cognitive effort, a well-founded and fully justified accurate state of joint attention (cf. Spaulding, 2016).

However, if we are interested in explaining how joint attention supports further coordinated activities, is this probabilistic, possibly fallible, notion of joint attention enough? In the next section I argue that it is.

## 3.3. Normative implications

The proposed account, I suggest, shifts the epistemic normative question from explaining how individuals can attain a justified and well-founded relation of mutual awareness, to explaining how individuals abstain from defaulting to possibly erroneous assumptions of mutual awareness. This approach concerns the mental and psychological states of each individual alone, so that all we can say, regarding A, is that A attends to *x*, and that A takes herself to be attending to *x* together with B.

The normative problem A faces is to *avoid* defaulting to her prior assumption that she jointly attends to *x* with B, when that default should not be made — that is, in cases where B does not share her attention. There are two basic ways in which A's having a psychological state of assuming mutually shared attention to *x* with B is not rationally justified: either none of conditions (B1) and (B2) hold; or condition (B1) holds but (B2) doesn't:

(B1). B attends to *x*.

(B2). B takes herself to be attending to *x* together with A.

Since A's psychological state of sharing attention to *x* with B does not need to include conditions (B1) and (B2), its rational justification is external to A's psychology. Her psychological state can be merely based on the following implicit reasoning: "I have a certain relation to *x*, and since you and I are so similar, (I assume) you have it too." Nothing requires A to reason any further. If she stops here, she takes herself to jointly attend to *x* with B (as a matter of her *psychology*, that is). She may, of course, be epistemically wrong and unjustified. This is one possible stage of epistemic failure. To override defaulting to assumptions of mutual awareness, however, A will have to face up the computational challenge involved in joint attention and estimate the following probabilities, previously mentioned: (P1) the probability that A herself is attending to *x* (here assumed to be not cognitively demanding); (P2) the probability that B does not attend *x*, and (P3) the probability that B's attention is not being shared (cf. Siposova & Carpenter, 2019).

These estimates constitute further psychological factors that modulate the rationality of her assumption of mutual awareness (or her withholding the assumption). Depending on her prior experience in domains involving objects such as *x* and people such as B, A may be more or less sensitive to relevant information for estimating these probabilities. According to her estimates, then, A will reject her assumption of mutual awareness and conclude that there is no joint attention with B, or A will keep her assumption. The accuracy of her estimates, however, is a matter of degree, and cannot put A in a fully justified and rational state of strict normative joint attention, since these estimates will never fully encompass, from the psychological stance of A alone, the *truths* of conditions (B1) and (B2). The truth of these two conditions is external to A, and therefore beyond the ken of her strictly psychological standpoint. Paraphrasing Herbert Clark (1996, p. 96), a strictly rational state of *normative* joint attention, i.e. one that is itself fully well-founded and justified, can only be held by an omniscient being. The rationality of A's assumption of mutual awareness is, on this proposal, a matter of degree.

One worry is that, since A's and B's estimates are a matter of degree and strictly external to each other, my account results in a radical individualist position that ignores the interactive relation between A and B, and the mental scaffolding that such interaction provides. In other words, the "jointness" of joint attention is taken away. It is correct to note that, on my proposal, A's and B's estimates are strictly external to each other, to the extent that A could take herself to be engaged in joint attention with, say, a non-animate object (as several experiments on joint attention seem to do, when they trick participants to believe they interact with others, when in reality they are interacting with an algorithm). But this doesn't imply, on one hand, that there is there is no psychological to-and-fro between A and B at all. It does imply, on the other hand, that such to-and-froing is not unproblematically or irrefutably transparent; but rather it is always mediated by behavioural cues and implicit, potentially fallible, assumptions and estimates regarding these cues, which in turn may lead to further behaviours and assumptions from A and B in such a way that they feed off each other. Establishing and maintaining joint attention is a dynamic process that allows for eventual mistakes and corrections. For

example, A's initial erroneous assumption of joint attention with B to *x* may make her engage in behaviours (smiles, pointing, vocalisations about *x*), which then form cues to B to assume, in her turn, that they jointly attend to *x*.

Another possible worry with this proposal is that it leads to the conclusion that a fully rationally justified state of joint attention is not achievable. If so, the worry goes, this would severely impair the functional role of joint attention as a rational base for further joint actions, joint beliefs, and for yielding shared knowledge of our environment (Campbell, 2018; Seemann, 2019; Wilby, 2023). Although the answer will ultimately depend on how strict we make the notion of epistemic rationality, I do not see compelling reasons to conclude that the role of joint attention in social coordination is thereby impaired or unexplained. There is no a priori need to assume that the functional role of joint attention is to provide *logically* irrefutable grounds allowing individuals to engage rationally in joint endeavours concerning the jointly attended object. This normative assumption not only ignores psychological plausibility, but it may also leave aside aspects of strategic rationality at play in joint attentional scenarios. William James already noted that "the logic of belief and knowledge" is too abstract to tackle many epistemic situations. We cannot simultaneously attempt to believe as many truths as possible and as few falsehoods as possible. According to James, we need to implicitly or explicitly weigh the value of avoiding false positives against false negatives (James, 1956; cf. Van Fraassen, 2002, p. 88).

In many situations, we don't need a normatively tight state of joint attention to engage in coordinated actions. Rather, different contexts and goals may require that we implicitly gauge whether to invest cognitive effort to override the assumption of joint attention. In a football game, for example, if I assume joint attention with an opponent toward the ball and get it wrong, I may not incur any significant costs, given my practical interests (and I may never realise I got it wrong). But if I don't assume joint attention, and should have (e.g., B1 and B2 are true, but A still fails to assume joint attention), the consequences could be drastic. Here, a false negative is more pernicious than a false positive. The coordinated-attack scenario is an extreme case in the opposite direction: if I assume joint attention and get it wrong, then we both stand to incur high costs. But if I don't assume joint attention, and should have, the costs are negligible. Here, a false positive is more pernicious than a false negative. It would seem rational for an agent to be sensitive to the payoff structure of a particular situation. A notion of epistemic justification in joint attention based on strict objective accuracy cannot capture such strategic aspects. A psychological state of being attending to an object or event together with another person may thus serve a functional role in action, without that state itself being necessarily accurate or irrefutably justified. More generally, a good enough assumption of joint attention, even if not completely logically justified, is often functionally strategic and sufficient for grounding further social coordination (cf. Stein, 1996; Todd & Gigerenzer, 2012). As Spaulding (2016) suggests, to successfully engage in social coordination and interaction, beings like us with limited cognitive resources will often use their mindreading abilities, including those deployed in joint attention, with the aim of being efficient rather than fully and justifiably accurate.

This suggests that the strict normative notion of joint attention — where the state of mutual awareness in joint attention must be fully rational and justified — may be eventually dispensable if we are interested in explaining how joint attention supports further joint action and coordination. After all, the clash between reductive and non-reductive approaches to joint attention has remained unsettled because both approaches start their analyses with such strict normative notion of openness in joint attention. A notion of joint attention based on the fallible

assumptions and estimates of each individual, even if not fully and logically well-founded, may well provide satisfactory functional grounds for joint action and shared knowledge in most situations. Importantly, this conclusion holds even when doing away with the second tier of my proposal. Specifically, the notion that joint attention is often implicitly assumed in a heuristic-like step is not strictly necessary for this point. As long as joint attention is supported by the (possibly) fallible and biased estimations of probabilities (P1) to (P3), positing the existence of a fully justified state of joint attention may become explanatorily expendable.

At a minimum, theoretical analyses of joint attention should endeavour to distinguish their explananda, as the project of understanding joint attention as a psychological phenomenon may be different in kind from the project of providing constitutive conditions of a normatively well-founded and strictly rational state of joint attention.

# Conclusion

People effortlessly engage with others in activities that require attending together to some object or event. Engaging in joint attention provides a basis for further coordinated activities, from joint actions to shared beliefs and communication. To address this function, it is widely held that joint attention is essentially "out in the open". Going beyond the metaphor of openness, however, requires an account of the mutual awareness that underlies joint attention. Some current accounts, I have argued, fail to distinguish between two distinct explanatory aims. One aim is normative: what conditions must hold to have a well-founded and rationally justified state of joint attention? Another is cognitive: what cognitive capacities and mental processes or understanding are involved in joint attention? Both reductive iterative and non-reductive relational accounts of the openness in joint attention are primarily concerned with the normative aim, and their limitations arise when they are taken to directly address the cognitive aim.

Taking a different methodological strategy, I suggest that engaging in joint attention involves the imperfect estimation of several uncertainties, where the openness in joint attention is not something that co-attenders must arrive at, but is often implicitly assumed. On this hypothesis, given the right sort of cues, people will tend to assume, often without any conscious reflection, that they are attending to some object or event together with someone else. On this probabilistic proposal, however, an individual cannot ever be fully aware of having rationally justified joint attention toward a common object with another agent, nor can they ever be fully aware of all the factors that make their state of joint attention rationally justified (cf. Sperber & Wilson, 1995, pp. 19–20). They can, at best, form more or less accurate estimates of these factors. These estimates may be sufficient to account for the functional role of joint attention in supporting social behaviours and joint actions.

# Acknowledgements

# References

Akkoyunlu, E. A., Ekanadham, K., Huber, R. V., Akkoyunlu, E. A., Ekanadham, K., & Huber, R. V. (1975). Some constraints and tradeoffs in the design of network communications. *Proceedings of the Fifth ACM Symposium on Operating Systems Principles*, *9*, 67–74.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*(4), 1278–1289.

Battich, L., Fairhurst, M., & Deroy, O. (2020). Coordinating attention requires coordinated senses. *Psychonomic Bulletin & Review*, *27*(6), 1126–1138. https://doi.org/10.3758/s13423-020-01766-z

Battich, L., & Geurts, B. (2021). Joint attention and perceptual experience. *Synthese*, *198*(9), 8809–8822. https://doi.org/10.1007/s11229-020-02602-6

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science*, *18*(5), 382–386. https://doi.org/10.1111/j.1467-9280.2007.01909.x

Brinck, I., & Gärdenfors, P. (2003). Co-operation and communication in apes and humans. *Mind and Language*, *18*(5), 484–501. https://doi.org/10.1111/1468-0017.00239

Bruner, J. S. (1995). From joint attention to the meeting of minds. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 1–14). Lawrence Erlbaum.

Burge, T. (2011). Self and self-understanding. *The Journal of Philosophy*, *108*(6/7), 287–383.

Calabi, C. (2008). Winks, sighs and smiles? Joint attention, common knowledge and ephemeral groups. In H. B. Schmid, K. Schulte-Ostermann, & N. Psarros (Eds.), *Concepts of sharedness: Essays on collective intentionality* (pp. 41–58). De Gruyter.

Campbell, J. (2002). *Reference and consciousness*. Oxford University Press.

Campbell, J. (2005). Joint attention and common knowledge. In N. M. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds. Issues in philosophy and psychology* (pp. 287–297). Oxford University Press.

Campbell, J. (2011). An object-dependent perspective on joint attention. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 415–430). MIT Press.

Campbell, J. (2018). Joint attention. In M. Jankovic & K. Ludwig (Eds.), *The Routledge handbook of collective intentionality* (pp. 115–129). Routledge.

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. 159–182). MIT Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i–174. https://doi.org/10.2307/1166214

Choi, Y., Mou, Y., & Luo, Y. (2018). How do 3-month-old infants attribute preferences to a human agent? *Journal of Experimental Child Psychology*, *172*, 96–106. https://doi.org/10.1016/j.jecp.2018.03.004

Clark, H. H. (1992). *Arenas of language use*. The University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Eilan, N. (2005). Joint attention, communication, and mind. In N. Elian, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds. Issues in philosophy and psychology* (pp. 1–33). Oxford University Press.

Eilan, N. (2015). *Joint attention and the second person (draft).* https://warwick.ac.uk/fac/soc/philosophy/people/eilan/jaspup.pdf

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, *40*(6), 760–768. https://doi.org/10.1016/j.jesp.2004.02.002

Farrar, B. G., & Ostojić, L. (2018). Does social distance modulate adults' egocentric biases when reasoning about false beliefs? *PLOS ONE*, *13*(6), e0198616. https://doi.org/10.1371/journal.pone.0198616

Gallagher, S. (2010). Movement and emotion in joint attention. In S. Flach, D. Margulies, & J. Söffner (Eds.), *Habitus in habitat I: Emotion and motion* (pp. 41–54). Peter Lang.

Gallagher, S. (2011). Interactive coordination in joint attention. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 293–305). MIT Press.

Geurts, B. (2019a). Communication as commitment sharing: Speech acts, implicatures, common ground. *Theoretical Linguistics*, *45*(1–2), 1–30. https://doi.org/10.1515/tl-2019-0001

Geurts, B. (2019b). Commitments continued. *Theoretical Linguistics*, *45*(1–2), 111–125. https://doi.org/10.1515/tl-2019-0009

Grosse Wiesmann, C., & Southgate, V. (2021). Early theory of mind development: Are infants inherently altercentric? In M. Gilead & K. N. Ochsner (Eds.), *The neural basis of mentalizing* (pp. 49–66). Springer International Publishing.

Haddock, A., & Macpherson, F. (2008). Introduction: Varieties of disjunctivism. In A. Haddock & F. Macpherson (Eds.), *Disjunctivism: Perception, action, knowledge* (pp. 1–32). Oxford University Press.

Hobson, P., & Hobson, J. (2011). Joint attention or joint engagement? Insights from autism. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 115–136). MIT Press.

James, W. (1956). The will to believe. In *The will to believe and human immortality* (pp. 1–31). Dover Publications.

Kim, E. Y., & Song, H. (2015). Six-month-olds actively predict others' goal-directed actions. *Cognitive Development*, *33*, 1–13. https://doi.org/10.1016/j.cogdev.2014.09.003

León, F. (2021). Joint attention without recursive mindreading: On the role of second-person engagement. *Philosophical Psychology*, *34*(4), 550–580. https://doi.org/10.1080/09515089.2021.1917533

Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press.

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, *12*(1), 142–149. https://doi.org/10.1111/j.1467-7687.2008.00741.x

Liddle, B., & Nettle, D. (2006). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology*, *4*(3), 231–244. https://doi.org/10.1556/JCEP.4.2006.3-4.3

McDowell, J. (1982). Criteria, defeasibility, and knowledge, *Proceedings of the British Academy*, *68*, 455–479.

McDowell, J. (2008). The disjunctive conception of experience as material for a transcendental argument. In A. Haddock & F. Macpherson (Eds.), *Disjunctivism: Perception, action, knowledge* (pp. 376–389). Oxford University Press.

Mole, C. (2017). Attention. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2017/entries/attention/

Moll, H., & Meltzoff, A. N. (2011). Joint attention as the fundamental basis of understanding perspectives. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 393–414). MIT Press.

Moll, H., Pueschel, E., Ni, Q., & Little, A. (2021). Sharing experiences in infancy: From primary intersubjectivity to shared intentionality. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.667679

Moll, H. (2024). What we do and don't know about joint attention. *Topoi, 43*(2), 247–258. https://doi.org/10.1007/s11245-023-09961-y

Moore, C., & Dunham, P. J. (1995). Current themes in research of joint attention. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 15–28). Lawrence Erlbaum.

Mundy, P. (2018). A review of joint attention and social-cognitive brain systems in typical development and autism spectrum disorder. *European Journal of Neuroscience, 47*(6), 497–514. https://doi.org/10.1111/ejn.13720

Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks*, *23*(8), 985–997. https://doi.org/10.1016/j.neunet.2010.08.009

Peacocke, C. (2005). Joint attention: Its nature, reflexivity, and relation to common knowledge. In N. M. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds. Issues in philosophy and psychology* (pp. 298–324). Oxford University Press.

Reddy, V. (2010). *How infants know minds*. Harvard University Press.

Reddy, V. (2011). A gaze at grips with me. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 137–158). MIT Press.

Roessler, J. (2005). Joint attention and the problem of other minds. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds. Issues in philosophy and psychology* (pp. 230–259). Oxford University Press.

Scarafone, A. (2024). Joint attention: Normativity and sensory modalities. *Topoi, 43*(2), 283–294. https://doi.org/10.1007/s11245-023-10004-9

Schiffer, S. R. (1972). *Meaning*. Clarendon Press.

Schmitz, M. (2014). Joint attention and understanding others. *Synthesis Philosophica*, *29*(2), 235–251.

Seemann, A. (2011). Introduction. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 1–17). MIT Press.

Seemann, A. (2019). *The shared world: Perceptual common knowledge, demonstrative communication, and social space*. MIT Press.

Seemann, A. (2024). Joint attention as the base of common knowledge and collective Intentionality. *Topoi, 43*(2), 259–270. https://doi.org/10.1007/s11245-024-10011-4

Siposova, B., & Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition, 189*, 260–274. https://doi.org/10.1016/j.cognition.2019.03.019

Scaife, M., & Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature, 253*, 265–266.

Spaulding, S. (2016). Mind Misreading. *Philosophical Issues*, *26*(1), 422–440. https://doi.org/10.1111/phis.12070

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (Second). Blackwell.

Southgate, V. (2013). Early manifestations of mindreading. In S. Baron-Cohen, H. Tager-Flusberg, & M. V. Lombardo (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience* (3rd ed., pp. 3–18). Oxford University Press.

Southgate, V. (2020). Are infants altercentric? The other and the self in early social cognition. *Psychological Review, 127*(4), 505–523. https://doi.org/10.1037/rev0000182

Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science.* Clarendon Press.

Ting, F., He, Z., & Baillargeon, R. (2021). Five-month-old infants attribute inferences based on general knowledge to agents. *Journal of Experimental Child Psychology, 208*, 105126. https://doi.org/10.1016/j.jecp.2021.105126

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, *107*(4), 657–676. https://doi.org/10.1037/a0037037

Todd, P. M., & Gigerenzer, G. (2012). What is ecological rationality? In P. M. Todd & G. Gigerenzer (Eds.), *Ecological rationality: Intelligence in the world* (pp. 3–30). Oxford University Press.

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). Lawrence Erlbaum.

Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, *28*(1), 509–529. https://doi.org/10.1146/annurev.anthro.28.1.509

Tomasello, M. (2003). *Constructing a language.* Harvard University Press.

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences*, *115*(34), 8491–8498. https://doi.org/10.1073/PNAS.1804761115

Tomasello, M. (2019). *Becoming human: A theory of ontogeny.* Harvard University Press.

Trevarthen, C. (2011). What is it like to be a person who knows nothing? Defining the active intersubjective mind of a newborn human being. *Infant and Child Development*, *20*(1), 119–135. https://doi.org/10.1002/icd.689

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry*, *42*, 3–48. https://doi.org/10.1111/1469-7610.00701

Van Fraassen, B. C. (2002). *The empirical stance.* Princeton University Press.

Vincini, S. (2024). Joint attention, relationalism, and individuation. *Philosophical Psychology*, *38*(5), 2024–2049. https://doi.org/10.1080/09515089.2024.2318421

Watzl, S. (2012). Review of Axel Seemann (Ed.) "Joint attention: New developments in psychology, philosophy of mind, and social neuroscience". *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/reviews/joint-attention-new-developments-in-psychology-philosophy-of-mind-and-social-neuroscience/

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wilby, M. (2010). The simplicity of mutual knowledge. *Philosophical Explorations*, *13*(2), 83–100. https://doi.org/10.1080/13869791003759963

Wilby, M. (2023). The form and function of joint attention within joint action. *Philosophical Psychology, 36*(1), 134–161. https://doi.org/10.1080/09515089.2022.2039384

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Zawidzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*. MIT Press.