



Joint Attention: Normativity and Sensory Modalities

Antonio Scarafone¹

Accepted: 21 December 2023 / Published online: 16 January 2024
© The Author(s) 2024

Abstract

Joint attention is typically conceptualized as a robust psychological phenomenon. In philosophy, this apparently innocuous assumption leads to the problem of accounting for the “openness” of joint attention. In psychology, it leads to the problem of justifying alternative operationalizations of joint attention, since there does not seem to be much which is psychologically uniform across different joint attentional engagements. Contrary to the received wisdom, I argue that joint attention is a social relationship which normatively regulates the attentional states of two or more individuals. This social account of joint attention leans on Bart Geurts’ view of communication as commitment sharing. Its promises are: (i) to explain the role of joint attention in wider joint activities, including communicative interactions; (ii) to account for how playing this role requires individuals to deploy different psychological resources on different occasions; and (iii) to identify the rationale behind alternative operationalizations of joint attention.

Keywords Joint attention · Communication · Commitment · Sensory modalities · Autism

1 The Problem

Joint attention is fundamental to our social lives. Buying bread from a baker, dancing, watching a movie together, having a chat about the weather, and demonstrating how to play a piece of music, are all activities which require us to jointly attend to things in our environment, be they objects, events, or, perhaps, thoughts. Joint attention is also widely regarded as a key steppingstone in child development (see, e.g., Tomasello 2019). Children who do not engage in joint attention in typical ways, often do experience serious difficulties in acquiring a language and in coping with the social world around them (see, e.g., Griffin and Dennett 2009).

Since Jerome Bruner (1974) introduced the technical notion of joint attention in developmental psychology, much ink has been spilt by philosophers and cognitive scientists to understand what we talk about when we talk about joint attention. It is typically assumed that joint

attention is a robust psychological phenomenon. For philosophers, a central concern has been to account for the difference between each of us attending to something individually, and the two of us attending to it jointly (see, e.g., Eilan et al. 2005, Battich and Geurts 2021). Looking out of my window, I am attending to the cows in the field. Looking out of your window, you are attending to the same cows. However, we cannot see each other from our respective windows, and we are not in any way aware of each other’s presence. Hence, there does not seem to be any joint attention. Intuitively, it seems that for our attentional states to constitute a joint attentional engagement, we must at least be aware, or know, that we are attending to the same thing. Let us assume, then, that we are out and about and that we can see each other attending to the cows. This will not be enough. Assume that this is a very funny pasture, interspersed with big one-way mirrors. Between us, there is a thick pane of glass which we mistakenly believe to be one such mirror.¹ In this situation, neither of us knows that each knows that the other is attending to the cows. Again intuitively, we still fall short of jointly attending to the cows.

✉ Antonio Scarafone
a.scarafone.philpsych@gmail.com

¹ Faculty of Philosophy, Theory of Science and Religious Studies Chair of Philosophy of Mind, LMU Munich, Geschwister-Scholl-Platz 1, 80359 Munich, Germany

¹ This example is adapted from Peacocke (2005). More examples with higher-order iterations can be found, for instance, in Campbell (2005).

The intuition is that for something to count as a genuine joint attentional engagement, it must be ‘wholly overt’ between us that we are both attending to the same thing, and the temptation is to paraphrase this metaphor in terms of iterations of knowledge states (*a* knows that *b* knows that *a* knows, etc., that *p*). As it is often argued, the problem is that no matter how many layers of knowledge one adds, there does not seem to be any safe resting spot before having achieved joint attention. And one feels compelled to postulate that the iterations of knowledge states must be infinitely many:

$$\begin{aligned} &K_a p \text{ and } K_b p. \\ &K_a K_b p \text{ and } K_b K_a p. \\ &K_a K_b K_a p \text{ and } K_b K_a K_b p. \\ &\dots \end{aligned}$$

Since it is assumed that joint attention is a robust psychological phenomenon, the problem becomes to square the intuitive ‘overt’ or ‘openness’ of joint attention, spelt out in terms of mutual knowledge, with the kind of awareness or knowledge that psychologically limited beings, including young children, can have. At this juncture, either the concern regarding psychological plausibility is shown to be spurious (Battich and Geurts 2021), or openness must be cashed out otherwise. For instance, it has been proposed to conceptualize joint attention in terms of a dispositional open-ended awareness (Peacocke 2005) or as a primitive perceptual relation holding between an object and two co-attenders (Campbell 2005).²

Despite the intuitive appeal of the metaphor of openness, it is not clear what these analyses of joint attention, understood as attempts to paraphrase the metaphor, are answerable to. The notion of joint attention was introduced in psychological theorizing as a technical one, and it is assumed that it denotes a psychological phenomenon. Therefore, the need to account for its ‘openness’ cannot be motivated as a piece of ordinary language philosophy. On the other hand, it is not easy to see how spelling out the metaphor in infinitary terms may respond to the need of explaining or systematizing empirical findings.³

So, normative accounts of joint attention, which are supposed to spell out the conditions that something ought to meet to count as a joint attentional engagement, seem to be caught up in a metaphor. As noticed by Wilby (2023), the

problem for psychologists is that, without a viable normative account, it is not easy to operationalize the very notion of joint attention consistently. To illustrate: capacities for joint attention are often operationalized as establishing eye contact and alternating gaze on the target object (Butterworth 1995, Laidlaw et al. 2016). However, participating in joint attentional engagements requires, and allows for, recruiting various psychological and sensory resources on different occasions (Battich et al. 2020). For instance, it is perfectly possible to jointly attend to something which is not visible, like a sound or a taste, and blind individuals can well participate in joint attentional engagements. To account for these and many other possibilities, as well as to draw sound conclusions about the psychology of joint attenders, it is necessary to make explicit the rationale behind alternative operationalizations of joint attention.

This paper is not another attempt to paraphrase the metaphor of openness, but to dispense with it. In Sect. 2, I will make a proposal concerning what an account of joint attention ought to explain. The main claim will be that an account of joint attention ought to explain what difference it makes to attend to something jointly, as opposed to individually, for the purposes of coordinating actions. I will argue that this difference can hardly be captured in solely psychological terms, partly because the psychological requirements vary too widely on occasions. To make my case, I will present a series of examples showing that, on different occasions of our joint attending, we are required to deploy different psychological resources, and what these are depends, crucially, on the wider joint activity in which the joint attentional engagement is embedded. In Sect. 3, I will detail my positive proposal: joint attention is a social relationship which normatively constrains the attentional states of two or more individuals.⁴ This account is based on the independently motivated, commitment-based conception of communication defended by Geurts (2019a, b). In a nutshell, the idea is that the difference between attending to something jointly as opposed to individually, is a difference in the commitments which are common ground for us. These commitments create distinctive affordances and constraints for action coordination. I will argue that this account makes it possible to individuate the rationale behind alternative operationalizations of joint attention, and it explains why, on different occasions of our joint attending, we must deploy different psychological resources. Finally, in Sect. 4 I will illustrate the benefits of this account with respect to scientific and clinical practice.

² As anticipated by Peacocke (2005) and then argued by Battich and Geurts (2021), it is hard to make sense of the notion of co-attender without making it collapse on the notion of mutual knowledge. I agree with the other objections raised by Battich and Geurts to Campbell’s relational account.

³ Some authors (e.g., Campbell 2005) motivate the problem by referring to coordination problems in game theory. I will come back to this point in another footnote in Sect. 3.

⁴ At least in spirit, there are important similarities between the account on offer here and Michael Wilby’s (2023). I provide a comparison in Sect. 3.

2 Identifying the Right *Explananda*: The Problem of the Many

An adequate account of joint attention must explain how our joint attentional engagements are based on, and sustain, coordination of action within wider joint activities.⁵ Two elements need to be accounted for: first, the fact that what each participant in a joint attentional engagement must do, differs on occasions. Second, the fact that having jointly attended to something enables individuals to further coordinate their actions and talk, again in various ways on different occasions. In both cases, the psychological resources which individuals need to, or can, rely upon vary according to the occasion. An account of joint attention is thus in the business of individuating the role of joint attention in action coordination, and to suggest how this role may constrain individuals' actions and psychological resources. In this section, I will make these *explananda* more concrete by means of examples.

We are hiking on the mountains when you point at an eagle and exclaim: "Look at that!". I'm supposed to at least try to identify what it is that you are pointing at, and I shall come back to you by sharing my thoughts and feelings about it.⁶ The goal of the enterprise is to align our attentional states, in this case on a visible target, and to respond to each other appropriately. Aligning our attention in this case may require, and be followed by, some communicative back and forth ("Where?!" "Close to the cross, to the right!" "I see it! Awesome!"). What one situation requires may well be impermissible in another. For instance, we might be biologists tasked with monitoring the fauna in a forest without disturbing it. Or, to vary the situation more dramatically, we might be at home watching a documentary together. Here, constantly alternating gaze while commenting on what is on the screen, is the exception rather than the rule, and in our subsequent conversation we will *presuppose* that we both watched the documentary with due attention.

Joint attention is often thought of as joint visual attention, and it is operationalized as the coordination of gaze on distal visible targets (see, e.g., Laidlaw et al. 2016). However, nothing in either the concept or the phenomenon imposes such a restriction. For instance, the common focus of attention might be a performance rather than an object, and we

might have to rely on sensory modalities other than sight. Prototypical face-to-face demonstrations, which feature joint attentional engagements,⁷ are a case in point. Imagine that you are demonstrating to me how to paint a miniature. There is something that you are doing attentively,⁸ namely, painting the miniature, and I am supposed to pay attention, primarily, to what you are doing, which here entails paying attention to your hands, how you hold the brush, which colours you use, and so on. Since you are an expert painter, beyond doing something attentively, you may well monitor what it is that I am looking at, to make sure that my attention is directed onto the relevant features of your performance. I might then want to try for myself. Now I must pay attention to what *I am doing*, and I must try to paint the miniature according to your teachings. I do not need to actively monitor what *you are looking at*, but I'd better remain sensitive to your feedback, thus *presupposing* that you are monitoring what I am doing with due attention.

So, the common focus of attention can be a performance rather than an object, and the sensory modalities relied upon by one party need not match those relied upon by the other. In some cases, moreover, sight might not be relied upon at all by either party. A blind individual may well be able to demonstrate how to model a clay pot, or how to play a piece of Chopin on the piano. A blind pupil may well learn from such demonstrations, and the teacher will often be able to monitor the pupil's attention and performance, though not in the visual modality. These and other similar variations in the relied-upon sensory modalities abound, and they stretch much beyond the context of demonstrations (Battich et al. 2020).

Joint attention is widely regarded as a key stepping-stone in child development (see, e.g., Tomasello 2019). Infant joint attentional engagements present dimensions of variation analogous to those of their adult counterparts. As Reddy (2008) convincingly argues, there are several forms of infant joint attention which do not consist in aligning

⁵ According to Gallagher (2011), joint attentional engagements are *the most minimal* form of joint activity. From the definition of joint activity that I adopt in this paper, it indeed follows that attentional engagements are themselves joint activities, but it does not follow that they are the most minimal ones. Also, I am not committed to other aspects of Gallagher's proposal.

⁶ The reader will notice that I employ a thoroughly normative vocabulary in presenting examples of joint attention. In the next section, I will make explicit why I think employing a normative vocabulary is not only apt but sometimes indispensable.

⁷ Demonstrations do not feature joint attentional engagements as a matter of conceptual necessity. One can record a demonstration, and somebody else can learn from it. Arguably, in these cases there is no joint attention. But these cases are derivative, and not only from a technological point of view. To learn from a recorded demonstration, one must be disposed to correct one's own performance according to the standard set by the demonstration, but without relying on the demonstrator's feedback. Learning from recorded demonstrations might thus require more cognitive maturity. Thanks to Ophelia Deroy for having pushed me to clarify this point.

⁸ A theory of joint attention should not prejudice the complementary analysis of attention. Though for most of the examples I will survey, a theory of attention as selection for action (Wu 2014) will readily do, for certain aspects of demonstrations, an adverbial theory (Mole 2011) might be more illuminating, as the phrasing I chose suggests. Thanks to Anpeng Liu and Marat Shardingaliev for having pushed me to clarify this point.

visual attention on a distal visual target. The common focus of attention might be, say, the infant herself, a body part, another person, an object-in-the-hand, an action performed by the infant, or an action performed by the adult. To illustrate: in showing off or clowning, the infant draws the adult's attention to her own performance of a certain *act*. We know that in these cases the infant is not merely calling the adult's attention to *herself*, because once she obtains the adult's attention, she keeps repeating the act. On the other hand, when infants are primarily calling for the adult's attention on themselves, and they obtain it, they stop calling, and this is indeed part of what makes the vocalization a *call*.

Infants show a remarkable facility in learning from demonstrations. A great deal of experimental studies (see, e.g., Csibra and Gergely 2009) have detailed our understanding of this form of learning, as well as what, in the eyes of the infant, marks an act as one of demonstrating.⁹ There is good experimental evidence that, when the demonstration (or the wider joint activity) revolves around an object which the adult is manipulating, the infant will monitor the adult's attention by attending, primarily, to the hands rather than the eyes. And more generally, joint attentional engagements are, for infants and adults alike, sustained by the formation of multisensory expectations (Battich et al. 2020).

Analogous points apply to both the learning of words and pointing. Joint attention is widely regarded as a key element in word learning.¹⁰ Let us assume that this is true. Vision might again be the privileged modality, but it is not necessary to rely on it, as testified by the fact that, if the right environmental conditions hold, blind children often acquire the vocabulary of their native language at a rate comparable to that of sighted children (Bloom 2002). Pointing is widely regarded as one of the royal ways for typically developing infants to establish and sustain joint attentional engagements.¹¹ Pointing seems to originate in development as a kind of ritualized touch (Carpendale & Carpendale 2010), and there is good experimental evidence that the expectations formed by the observer integrate information of a visual and tactile nature (O'Madagain et al. 2019).

Even when the point of the enterprise is to jointly attend to a distal visual target, and this bout of joint attention is established by means of pointing, what one must do to count as jointly attending to that object together with the other person, varies according to the situation. Assume, for instance,

⁹ For an illustration of some of the normative import of demonstrations, and a proposal on how this normative import might have come to light in our evolutionary past, see Weiss 2022.

¹⁰ The dust has not yet settled on this issue. For at least some important aspects of language acquisition, it might be enough to *witness* conversational exchanges embedded in wider joint activities without *participating* in joint attentional engagements (see, e.g., Kissine 2021).

¹¹ For a detailed account of infant pointing which complements the present account of joint attention, see Scarafone and Michael 2022.

that Aunt Marie is playing a hide-and-seek game with little Freddie. She hides a toy in a box, and he is supposed to find it. Little Freddie looks a bit puzzled, and so Aunt Marie helps him by pointing to the box where the toy is. Attending to the box, here, entails taking it as a location at which to look for the toy, which is something that 18-month-olds unproblematically and reliably do (see, e.g., Behne et al. 2005). If Aunt Marie and little Freddie were not playing a hide-and-seek game, then attending to the box might have consisted in doing something quite different. For instance, if a puppet comes out of the box and the infant is amused by it, the infant might excitedly point to the box and smile at the adult, without any sign of wanting the puppet for himself. Here, what it is for the adult to attend to the box may very well consist in looking at it and responding to the infant's amusement or surprise (see, e.g., Liszkowski et al. 2004).

On top of sensory modalities and actions, a further critical dimension of variation regards the prior knowledge that individuals must or can bring to joint attentional engagements, as well as what they can learn from participating in them. To make the point more vivid, I will consider the case of infant-adult joint attention but, *mutatis mutandis*, the same holds also for the adult-adult case. The corollary will be that, if joint attention is analyzed in terms of mutual knowledge, or in terms of other cognate epistemic/doxastic notions, then it is hard to explain the pedagogical import of joint attention.

To jointly attend to something, we must attend to the *same* thing. Satisfying this minimal requirement is not trivial. First, it is not obvious that infants parse out the surrounding environment in objects and their properties. Up to a point early in development, they might well rely on feature discrimination (Hildebrandt et al. 2020). If this is the case, then it remains to be explained what it is that the infant and the adult are (perhaps only allegedly) jointly attending to. Even under the assumption that infants do pick out objects and their properties, it still needs to be explained what it is, for the infant and the adult, to align their attention on the *same* object or on some of its properties. For instance, upon seeing a candle for the first time, the child does not yet know all that Daddy knows about candles. And for Daddy, candles are not quite the same magical sort of thing that they apparently are for the child.

Arguably, infants and adults come to align their attentional states on the same thing partly by interacting and correcting each other. To illustrate: the child attends to the demonstration and then strives to do the same thing. Criteria for sameness here may be broken down in terms of achieving a certain goal, reproducing certain acts, or doing something in a certain manner. A key ingredient to establish sameness in the eyes of the pupil, beyond perceived similarities in outcome and performance, is the teacher's

acceptance, rejection, or correction of what the pupil does. Young children are indeed sensitive to others' normative attitudes toward their own doings and, what is more, there is good experimental evidence that they also correct third parties when they act defiantly (see, e.g., Schmidt et al. 2019).

Establishing joint attention successfully contributes to the child's coming to know what counts as attending to the same thing. Arguably, children learn a great deal of what they know about people and actions, as well as objects and their properties, by doing things together with adults and other children, where this often requires participating in joint attentional engagements. If, on many occasions, *coming to know* what counts as the same is, in part, an outcome of having jointly attended to it, then it is hard to explain how joint attention could *necessarily* comprise *mutual knowledge* that the infant and the adult are attending to the same thing. In the next section, I will make a proposal for how to conceptualize learning via joint attention. Before concluding the present section, I briefly elaborate on some further reasons why I think mutual knowledge cannot be a necessary ingredient of joint attention.

First, for a and b to mutually know that p , both a and b must know that p , and thus they must have the conceptual resources which possessing this bit of propositional knowledge requires. For the reasons outlined above, coming to have the required conceptual resources will often be a result of, rather than a requirement for, having successfully established joint attention on a variety of different occasions. Second, for a and b to mutually know that p , both a and b must have normal perceptual and inferential capacities and assume that the other has these capacities too. These assumptions make it the case, or so it is argued, that, if such and such conditions hold, upon knowing that p and knowing that you know that p , I also know that you know that I know that p , and so on, *ad infinitum*. For current purposes,¹² the main problem is that assumptions of normality are not generally satisfied by the young child. Having normal perceptual and inferential capacities, and assuming that the other has them too, is an endpoint of development, not the start.

In sum, there can be considerable asymmetries regarding the knowledge that each joint attender can deploy or attain. So much so that accounting for joint attention in terms of mutual knowledge does not seem plausible. Though these asymmetries are deeper and more evident when considering infant-adult joint attention, they are arguably present in the adult-adult case too (for instance, I might not yet know what

I am supposed to attend to when the sailor is trying to teach me how to tie knots in those exotic ways).

What the examples briefly reviewed in this section are meant to illustrate is that in many and perfectly mundane joint attentional engagements, there are important variations in what each of us is required to do, or entitled to presuppose, in jointly attending to x . And there are corresponding variations in the knowledge, motivations, and sensory modalities that each participant must, or can, rely upon. These variations are not random: they seem to depend in systematic ways on the nature of the wider joint activity in which the bout of joint attention is embedded. Pairwise, having successfully established joint attention seems to play a systematic role in enabling us to further coordinate what we do and what we say. An account of joint attention should then illuminate its role in action coordination, and it should also make sense of how fulfilling this role, on different occasions, constrains in various ways the psychological resources which individuals can or must deploy.

3 Joint Attention as Commitment Sharing

Human activities, be they discursive or not, are pervasively normative (for detailed illustrations of this point, see Brandom 1994, Weiss 2022, Geurts, *forthcoming*). From playing a game of hide-and-seek to driving on the highway and managing a large corporation, we take up roles, and we update our normative statuses as our activities unfold. If I enter a bakery as a client, I am entitled to get the bread I want if it is on offer and I pay what it costs; on these conditions, the baker is committed to give me the bread. When I walk out with my baguette, I stop being a client, and if I bump into a random person on the street, I will not have with them the same commitments and entitlements that I had with the baker. Without normative talk, the social world would be very obscure. The relevant norms can be written in a code of law or verbally agreed upon, or they can be implicit in the conventions we follow. Different kinds of norms admit of different degrees of vagueness in their formulation, and they may have more rigid or more flexible boundaries for what counts as compliance or violation. In some cases, the norms might not be articulated at all, and in many cases, it might not be possible to specify the relevant body of norms exhaustively. What matters, however, is that there are practices, and that within these practices there are behaviours such as correcting, accepting, blaming, praising, punishing, and so on which, at least implicitly, set normative standards.

These practices and behavioural dispositions, more than the ability to articulate a rule, warrant normative talk in relation to joint attention, as the phrasing of the examples in the previous section suggests. The proposal elaborated in this

¹² I provide further arguments against mutual knowledge and other cognate notions in Scarafone 2023. As I explain in Sect. 3, my mutualistic notion of choice is Geurts' normative conception of common ground, spelt out in terms of shared commitments. For a more comprehensive and detailed discussion of these mutualistic notions, their explanatory roles, and the relationships between them, see Geurts, *under review*.

section is that joint attention is a social relationship which normatively constrains the attentional states of two or more individuals, according to the wider joint activity. To make this idea more precise, I will introduce a minimal notion of joint activity (Ludwig 2007), couple it with the commitment sharing view of communication (Geurts 2019a, b), and finally plug in my proposed characterization of joint attention. Once the whole machinery is assembled, I will highlight what I see as its main selling points.

Minimally, a joint activity can be thought of as an event featuring multiple participants (Ludwig 2007). According to this broad characterization, not every joint activity necessarily features a common goal or a shared intention, regardless of how these notions are understood. Also, this characterization does not have any implication regarding the psychology of the participants. It will be assumptions regarding specific kinds of events, and what is required to participate in them successfully or appropriately, that will ground hypotheses about the psychology of competent participants.

When it comes to regulating joint activities, in humans this is often done communicatively, and communication is itself something that we do together. This idea forms the background for much work in contemporary pragmatics and psycholinguistics (e.g., Clark 1996). Geurts (2019a) rephrases it with the slogan: *communication is coordinated action for action coordination*. According to Geurts, promising, requesting, asserting, giving an appointment, and so on, are all, primarily, things that we do together and that enable us to plan our activities effectively. If I promise you that I'll do the dishes, I must plan my activities so as to fulfil my promise, and in your own planning you can rely on the assumption that I will do the dishes. By the same token, if I ask you to do the dishes, and you grant my request, the onus of doing the dishes is on you, and I can do something else in the meantime.

The general idea is that speech acts create commitments and entitlements for both speakers and hearers. This enables us to manage our expectations and plan our activities, and thus to stably coordinate our actions over time.¹³ Commitments are modelled by Geurts (2019a) as ternary relationships between two individuals and a proposition: *a* is committed to *b* to act consistently with the truth of some proposition *p* (schematically: $C_{ab}p$). If I promise you that I'll do the dishes, and you accept this much, I become committed to you to act consistently with the proposition that I will do the dishes.¹⁴ In planning your activities, you are

thereby entitled to rely on the assumption that I will do the dishes. If you act on this assumption, but it turns out to be false, you are entitled to hold me responsible.

As socio-normative relationships, commitments can be in place even if one does not know or believe that they are. For instance, one might lack the conceptual resources to think about commitments, or to entertain a thought with the same content as that of the commitment. One might forget about a certain commitment or fail to see some of its consequences. Still, commitments have the potential to regulate the interaction. For instance, imagine that the adult says "Let's play Daxing! Here's how you play Daxing...", and then demonstrates to the young child how Daxing is played. Imagine, further, that somebody who has agreed to play the game does something which is not a legal move in the game, and so the child is entitled to correct them. 18 months olds do sometimes spontaneously correct deviant behaviour in these situations (Schmidt et al. 2019). Arguably, however, they do not yet know what commitments are. The same holds for adults: even though they may have the conceptual resources to think about commitments and a wide range of contents, on occasions they come to undertake commitments unknowingly, and those commitments do normatively regulate their interactions. For instance, I might sign up for a home equity line of credit, and even if I do not know what this is, I am in fact committed to repaying the loan, and I will be sanctioned if I don't.¹⁵

Geurts (2019a) proposes a generalization according to which not only promises, but speech acts of any kind, including assertions, questions, commands, acts of christening, etc., create commitments and entitlements for both speakers and hearers, to act in accordance with the truth of some proposition. So, for instance, if I ask you to do the dishes, and you grant my request, I commit myself to you to *you* doing the dishes. For both commissives (e.g., promises) and directives (e.g., requests), the content of the commitment specifies a goal, and either the speaker or the hearer (respectively) must see to it that the goal is achieved. Constatative speech acts (e.g., assertions) create commitments too, thereby constraining further sayings and doings, but in this case the content of the commitment typically does not specify a goal.

problem of which theory of meaning should sit alongside this conception of communication. The content will be determined by other commitments in the common ground, and inferential role semantics (especially as elaborated by Brandom 1994 and Drobnák 2022) is the most promising candidate passenger.

¹³ Bratman's work (e.g., 1987) is of fundamental importance to these discussions. Rather unfortunately, lack of space prevents me from detailing the fine relationships between Geurts' and Bratman's conceptions of commitment and action coordination.

¹⁴ Here, I will not delve into the problem of how the propositional content is determined (to the extent that it is), nor into the related

¹⁵ I think that, for children and adults alike, sharing commitments 'unknowingly' in this sense is not at all unusual, under the assumption that there is a practice, and that the individual can participate in that practice with some degree of competence. Much more should be said on this point. I elaborate on some of the related details in Scarafone 2023.

As social relationships, commitments cannot be in place unless it is accepted by both parties that they are. This point is of paramount importance, so it is worth illustrating it in some detail. If I say: “I’ll do the dishes tonight”, but you reply by saying “Don’t worry, I’ll take care of the dishes later”, then I have not promised you anything. My speech act remains a candidate promise, and this is why you will not be entitled to ask: “Why didn’t you do the dishes?!”. Conversely, if you accept my candidate promise, then I do become committed to you to doing the dishes. If you then start hindering my efforts, I am entitled to protest. Why? Because your acceptance of my promise is itself a kind of commitment that you have undertaken to me, $C_{ba}x$, and which has my promise ($C_{ab}p$) as its content:

$$C_{ba}C_{ab}p.$$

On the other hand, if I promised you that I will do the dishes, I must be prepared to acknowledge this much. If you then ask: “Why didn’t you do the dishes?!” I am not entitled to say: “I never said I would”. I am not entitled to this claim because, if I did promise, then I am committed to accepting this much. Here too, acceptance is a further commitment that I have undertaken to you ($C_{ab}x$), and which takes my promise ($C_{ab}p$) as its content:

$$C_{ab}C_{ab}p.$$

It bears emphasis, once again, that commitments are here thought of as social relationships, and acceptance, as a kind of commitment, is a social relationship too. Commitments are relational through and through. In this sense of ‘acceptance’, acceptance is not a psychological state, and one might be unaware of what one has accepted, or of its consequences. Acceptance is not an act either and so, even if it is often explicitly signaled, it needs not always be. Acceptance is entailed by the right kind of response (greeting in response to an act of greeting, answering a question, obeying an order, and so on) or, if minimal conditions for uptake hold, it is taken for granted. That said, in conversation acceptance is frequently signaled (e.g., nodding, saying ‘sure’, echoing, and so on). Within certain practices, certain commitments cannot come to hold unless there is a specific act of acceptance, as in signing a contract. Though, sadly, in this case too one might not be aware of what one has accepted (for a fuller discussion of these points and further examples, see Geurts 2019a, b).

So, if a commitment is in place, both parties accept it, and acceptance is itself a kind of commitment. Geurts (2019a) elegantly models these conditions with the following two rules of inference:

$$C_{ab}p \rightarrow C_{ba}C_{ab}p.$$

$$C_{ab}p \rightarrow C_{ab}C_{ab}p.$$

As we will see in a moment, these two rules generate the structure of common ground. The notion of common ground is at the heart of every theory of communication. Intuitively,

it is meant to capture how our communicative exchanges take place in, and constantly update, a context which is in an important sense shared. To communicate effectively, we may need to rely on conventions, identify the referents of pronouns, disambiguate expressions, respond appropriately to speech acts, draw implications and implicatures, relate what we say to the course of the conversation, take turns, and so on. To do all this, common ground is of the essence.

As Geurts explains (2019a), the rules governing acceptance generate the infinitary structure which is the fingerprint of common ground, and which is the same structure as that of mutual knowledge, mutual belief, and so on. To see how, consider that, if $C_{ab}p$ is accepted on both sides, it follows that:

$$C_{ab}C_{ab}p \text{ and } C_{ba}C_{ab}p.$$

$$C_{ab}C_{ba}C_{ab}p \text{ and } C_{ba}C_{ab}C_{ab}p.$$

$$C_{ab}C_{ba}C_{ab}C_{ab}p \text{ and } C_{ba}C_{ab}C_{ba}C_{ab}p.$$

....

The beauty of this framework is that, since there is a generalization according to which speech acts of any kind create commitments to act in accordance with the truth of some proposition, and for a commitment to be in place it must be accepted that it is, there is a general and unitary explanation of how felicitously addressed speech acts of any kind update, and are based on, common ground.

The final ingredient that we need is the notion of shared commitment, which Geurts (2019a) defines as follows:

$$C_{ab}p \text{ and } C_{ba}p.$$

The important upshot is that, if a commitment is shared, it is thereby common ground¹⁶ and, as we will see in a moment, this is what makes all the difference for action coordination.

The notion of commitment is a normative one, and this makes it plausible to claim that it is closed under entailment (if $C_{ab}p$ and p entails q , then $C_{ab}q$), an assumption which is far from trivial for knowledge or belief. Commitments can be shared without knowing or believing that they are shared. So, although we *can* reason about, and represent, the structure of common ground or some of its components in more or less accurate ways,¹⁷ this structure itself does not directly

¹⁶ The second rule of acceptance above is necessary to distinguish between acceptance and sharing, and this is key to account for disagreement (Geurts, manuscript). If a and b disagree on whether p , they do not share a commitment to p . However, if they are not talking past each other, they must at least share a commitment to one of them being committed to p ($C_{xy}C_{xy}p$ and $C_{yx}C_{xy}p$) and, presumably, they must also share a commitment to the other not being so committed ($C_{yx}\neg C_{yx}p$ and $C_{xy}\neg C_{yx}p$) or to the other being committed to the negation of the same proposition ($C_{yx}C_{yx}\neg p$ and $C_{xy}C_{yx}\neg p$).

¹⁷ I mentioned at the outset that some authors see the need to analyze common ground in infinitary terms to address coordination problems in game theory (e.g., Campbell 2005). Space prevents me from expanding on this point, but the commitment sharing view of common ground is especially well-suited for the study of these problems

mirror anybody's psychological processes, and so there is nothing problematic in its infinitary nature. This structure simply reflects the logic of commitment sharing. In a nutshell, Geurts' conception of common ground does justice to Lewis' (1969: 53) insight that the infinitary structure is 'a chain of implications', or of *admissible* inferences, and does not reflect anybody's psychological processes.

With all the main elements of the commitment-based account of communication in place, we can now plug in the candidate definition of joint attention:

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.

Let's see how this works. What each of us must do if we are to count as jointly attending to *x* depends on the occasion of our joint attending. This idea can now be made more precise by saying that, for each of us, the content of the commitment to attend to *x* is specified, to the extent that it is, by other commitments which we have in our common ground. These commitments are inferentially related to the commitment to attend to *x*, and they may specify, for instance, the goal of the wider joint activity (e.g., to learn how to model clay), our respective roles (e.g., demonstrating or attending to the performance), and changes in our normative statuses (e.g., becoming skilled). Correlatively, we can rely on the assumption that we have jointly attended to *x* to further coordinate our doings and sayings. This idea, too, can be made more precise by saying that, if we are jointly attending to *x*, it is in our common ground that we are, and we are thereby entitled to rely on this assumption in our subsequent interactions. So, for instance, sharing a commitment to attend to what I am doing entitles me, in the role of the demonstrator, to call back your attention if I notice that you are distracted. Pair-wise, having watched a movie together entitles each of us to refer to it during our subsequent conversation.

By default, commitments persist until they are fulfilled or reneged. If I am committed to you to attend to *x*, for how long do I have to keep attending to *x*? This, too, depends on what it is to attend to *x* on that occasion, as well as on what you and I are disposed to accept in the unfolding of

the interaction. If am demonstrating to you how to model a clay pot, the length of time for which you are committed to attend to my performance is, presumably, the duration of my performance. Not every bout of joint attention might be so neatly circumscribed. If you are excited about the eagle and invite me to attend to it with you, and I do look at the eagle for a sound 10 s and respond to you, presumably I'm in the clear. If we disagree about that, then we will question the boundaries of what counts as honouring the commitment to attend to the eagle on that occasion. Boundaries might not be exhaustively specified in advance. In principle, they could always be specified further, and they are set, at least in part, by what we are disposed to accept from each other.

The second conjunct in the definition, 'and they behave accordingly', is meant to rule out cases in which the commitment is shared, but one fails or does not even try to live up to it. If I tell you "Look at that!" and you say "Amazing!" but you do not see the eagle nor look in the direction where I point, we are not jointly attending to the eagle. Nevertheless, since we did share the commitment, in our subsequent interactions we are both entitled to rely on this assumption. At our peril, as it were.

In the literature, it has become common coin to distinguish between 'bottom-up' and 'top-down' joint attention, according to what prompts the joint attentional engagement. For instance, if each of us is absorbed in thoughts, but a goat appears in our living room and we look at each other, our bout of joint attention is 'bottom-up'. But if, for instance, we are a looking for a goat which disappeared, and I tell you "Look where it is! So cute!", while pointing to the roof, our subsequent bout of joint attention is top-down, because it is established in light of a goal which we have in our common ground. As far as joint attention goes, the commitment-based framework does not discriminate between goats which appear in living rooms and goats which get lost on roofs. Whether it is top-down or bottom-up, joint attention invariably entails sharing a commitment to attend to *x*, where, especially in the 'bottom-up' case, this commitment might just be 'acknowledged in practice', to say it *à la* Brandom. It bears emphasis that, even in this case, the commitment is forward looking. Assume that, by looking at each other with surprise, we acknowledge the presence of the goat. If you then say: 'Don't you find it unusual that that goat is dressed up like Santa Claus?', I am not entitled to respond: 'Which goat?!', because we did jointly attend to the goat.

In spirit, the current account is close to Michael Wilby's (2023), and so it will be useful to briefly compare the two. Wilby characterizes the main function of joint attention as follows:

(thanks to Jurgis Karpus for a helpful discussion on this point). It is perfectly possible to *represent* the infinitary structure of common ground and to reason about these assumptions (Karpus and Radzvilas 2021). Doing so is sometimes necessary to demonstrate what the point of equilibrium in a game is. However, if a given coordination problem were to require one to *carry out* every step of an infinitary reasoning, then there simply might not be any (finitary) solution to that problem. This distinction is easy to draw if common ground is a normative construct, but not if it is a psychological one (for a fuller illustration of the distinction between representing and having, with respect to common ground, see Geurts, [under review](#)).

JA has the normative function of enabling subjects to coordinate their actions in a way that would contribute to the rational justification of the execution of a joint action in accordance with a prior shared plan or shared intention. This is to say that JA has the function of providing agents with good reason for coordinating their actions at a particular time and place in pursuit of a particular goal.

Wilby 2023: 2.

Wilby explicitly restricts his attention to joint actions which comprise shared intentions. The account here on offer does not, because it relies on a minimal characterization of joint action (Ludwig 2007) as the participation of multiple individuals to the same event. Geurts' (2019a) notion of joint activity is similarly general. On the other hand, commitments do provide reasons for coordinating actions in certain ways (this is, after all, part of their job description), and they do so regardless of whether the joint activity features shared intentions. In this respect, the current proposal is more general than Wilby's.

There is a further point which is important to flag up, even if I cannot elaborate it here in detail. On Geurts' (2018) account, the normative import of intentions and beliefs is conceptualized in terms of private commitments, namely, commitments that one shares with oneself. Private commitments which specify goals take up the job of intentions, and those which do not, take up the job of beliefs. The addition of private commitments provides further resources to account for constraints of rationality, as well as for the very notion of shared intention (as, say, a certain constellation of social commitments to having certain private commitments, on both sides). Of course, these considerations ought to be further developed, but if they are on the right track, the current proposal arguably subsumes Wilby's in all relevant respects.¹⁸

One word about the metaphor of openness, on which so much ink has been spilt, is in order. This metaphor was meant to capture the distinction between attending to something individually and attending to it jointly, where this latter notion is supposed to be closely related to that of common ground. We can now draw the same distinction and dispense with the metaphor. On the present account, the difference is captured in terms of what difference it makes to attend to something jointly, as opposed to individually, for the sake of coordinating actions. This difference consists in there being

a social relationship between us which normatively constrains our psychological states. Certain commitments and entitlements to attend to things in our environment are in our common ground, and they would not be there if we were merely attending to those things individually. These shared commitments, in virtue of being in the common ground, create specific affordances for future sayings and doings, thus fostering action coordination.

Finally, it is worth making explicit how the commitment-based account scores in relation to learning from demonstrations, especially for the infant-adult case. It is not obvious to assume that the infant and the adult are in fact attending to the same object or performance. However, it is plausible to assume that they share a commitment to doing so, where this is evidenced, to the extent that it is, by their dispositions to behave normatively toward one another (Schmidt et al. 2019). If they strive to behave in accordance with the commitment, in the normal run of events they are likely to come to fulfill it too, and so their attention will be aligned with the adult's. In turn, this successful alignment can be, for infants, a road to knowing what it is that they are doing with the other person, what counts as doing the same thing, and so on. So, it is possible to make sense of the idea that infants learn from demonstrations, and hence via joint attention, without assuming that infants *know* in advance *that* they are attending to the same thing as the other person. Now, infants do not display dispositions to behave normatively from day one, and one might argue that they must develop a fair share of behavioural dispositions before having the normative ones. It would follow that, at this ancestral stage, there cannot be joint attention. True, but the infant-adult interaction will have some of the features of the later developing, genuinely joint attentional, interaction. The more ancestral interaction prepares the infant to take part in later occurring, genuinely joint attentional, engagements.¹⁹

4 The Psychology of Joint Attenders

The commitment-based account of joint attention is a normative account, in both attitude and subject matter. It tries to spell out which conditions ought to be met for two or more individuals to count as jointly attending to something, and it takes these conditions to be best formulated in terms of commitments. The commitment-based account is not a scientific theory of a social phenomenon, nor a theory of the psychology of joint attenders. Nevertheless, it provides a convenient conceptual framework to engage in these scientific activities. I will make my case by considering well-known difficulties experienced by autistic individuals in

¹⁸ Inspired by Campbell's work, Wilby then develops a relational account of JA which is supposed to explain how the function of JA is realized. Lack of space prevents me from discussing this point, but for the reasons provided in Sects. 1 and 2, and that will be further developed in Sect. 4, I am skeptical of any attempt at accounting for joint attention in solely perceptual or cognitive terms, no matter how sophisticated these accounts can be.

¹⁹ Thanks to Bart Geurts for having pushed me to clarify this point.

establishing joint attention, and less well-known findings and techniques which are suggestive of ways around those difficulties.

Avoiding eye-contact and refraining from establishing joint attention are widely held as key diagnostic hallmarks of autism in early childhood. These two things are often seen as one and the same which, as we will see, is unfortunate. In their auto-biographical writings, some autistic people have emphasized that establishing eye-contact is often, for them as adults, an emotionally overwhelming experience. A plausible conjecture (made by Reddy 2008) is that the same may be true also for autistic people who never come to acquire a language to describe their experiences. The tendency to avoid eye-contact may originate, at least in part, in a motivational or emotional asymmetry rather than in a cognitive or perceptual deficit. This initial asymmetry, conjoined with other factors, may well have cascading effects. For instance, avoiding eye-contact is likely to result in not looking directly at faces, which makes it harder to categorize and recognize expressions of emotions and, correlatively, to master the use of words employed to talk about emotions.

Several studies suggest that avoiding looking at the eyes of others stems from an emotional overload, a conjecture now known as ‘the eye-avoidance hypothesis’. In the longitudinal study by Jones and Klin (2013), it was found that infants later diagnosed with ASD started to manifest a gradual decline in their attentional orienting towards others’ eyes from the second to the sixth month of life, while up to that point they would be in the normal range. What this suggests is that infants later diagnosed with autism did not have difficulties in *detecting* the eyes of others as something salient.²⁰ It was also found that establishing eye-contact tends to put autistic children in a state of emotional distress, where this was evidenced by an abnormal hyperarousal of the amygdala.²¹ Though some uncertainty remains, several follow up studies which employed a variety of neuro-physiological measures have confirmed this hypothesis (for a recent review, see Stuart et al. 2023).²²

Failures of joint attention may then be due to emotional and motivational factors, and not only to perceptual or cognitive ones. The commitment-based framework is not in the business of providing scientific hypotheses, but it

²⁰ For data about typically developing infants and their tendencies to orient towards the eyes of others, see Csibra 2010.

²¹ The question under discussion in these studies was whether autistic children avoid establishing eye-contact because they are poorly motivated to attend to eyes or because it is distressing for them do so. Though possible, it might not be easy to experimentally discriminate between these two, competing hypothesis without employing neuro-physiological measures.

²² Thanks to Will McNeill for having initially brought these findings to my attention, and for a helpful discussion on their import for the present work.

makes it easier to formulate some. Indeed, sharing a commitment to attend to something on an occasion, and behaving accordingly, may require one to recruit emotional and motivational resources, on top of cognitive and perceptual ones. Depending on which of these components are either missing or too strong, there are different reasons why one avoids engaging in joint attention, or why one fails to follow through once one has accepted to participate. Different kinds of failures could be revealed in different patterns of behaviour, and they might require, or allow for, different kinds of remedies.

If the obstacle is an emotional one, insisting on *habituating* autistic children to eye-contact risks being both pointless and harmful. As we have seen in the previous sections, although establishing eye-contact might be part of what it is to participate in joint attentional engagements in prototypical situations, it is by no means a necessary component of joint attention as such. And if one modality is not available, at least on some occasions another modality can be used in its wake. It is then worth exploring different ways of making contact which do not start out with the eyes. This is what Phoebe Caldwell and, independently, Jaqueline Nadel and colleagues have been doing for some time now. Vasudevi Reddy gives a vivid, and touching, depiction of Caldwell’s technique which is worth quoting in full:

Gabriel was a young man with extremely severe autistic spectrum disorders, who spent most of the day in concentrated flapping and noise-making, and had essentially been given up on by scores of specialists. They simply could not make psychological contact with him. Phoebe’s attempts to communicate with Gabriel – all on film – lasted three days. Within 20 min on the very first day, imitating his small repetitive flapping movements, she tuned in to one aspect of them – his sensory interest in touching his left hand with any object that he was flapping – and she did the same herself on her own hand. Gabriel moved almost magically from the closed self-absorbed focus he habitually showed to a quieter, more outward focus, casting occasional looks at her hands while they both flapped in turn. Within two days (actually 5 h of working time) Phoebe and Gabriel were spending long moments of contact and silent mutual gaze.

Reddy 2008: 44.

It is worth noticing that, before having established eye-contact, when Caldwell attunes her own movements to Gabriel’s, and Gabriel attends to those movements, *they are establishing a joint attentional engagement*, which is in many ways analogous to dancing together. So, there are ways of making contact which may sidestep, and on some

occasions at least temporarily remove, some of the emotional obstacles on the way of establishing eye contact.

Adopting a commitment-based account is convenient for making sense of pre-conditions for establishing joint attention, formulate hypotheses about different behavioural patterns, and investigate which subsidiary resources can be deployed if the obvious ones are not there. The commitment-based account is not a scientific theory and does not by itself suggest any specific hypothesis. Nevertheless, it is a convenient framework because it provides criteria to classify certain behavioural patterns as instances of joint attention, and it accommodates the fact that the psychological resources may vary on occasions.

It is of course possible that alternative accounts of joint attention, which similarly focus on behavioural patterns but manage not to employ a normative vocabulary, would do just as well. For current purposes, the point is that alternative accounts which provide solely psychological criteria for individuating joint attentional behaviours, do not do just as well. If joint attention is said to be a perceptual phenomenon (e.g., Campbell 2005) or a cognitive one (e.g., Peacocke 2005), it is admittedly hard to identify joint attentional behaviours *independently* of their (alleged) psychological underpinnings. And these underpinnings can in fact vary. So, these accounts risk being misleading for the purposes of investigating failures of joint attention, as well as the psychology which does underlie joint attenders' behaviours on different occasions.

5 Conclusion

On the current proposal, joint attention is a social relationship which normatively constrains individuals' actions and psychological resources, according to the demands of the wider joint activity. This idea can be made precise by adopting the commitment-based conception of communication, and by plugging in a definition of joint attention as sharing commitments to attend to something and behaving accordingly. This conception of joint attention can systematically account for the role and place of joint attentional engagements within wider joint activities, including our communicative exchanges. And it can also account for the varying demands and opportunities that joint attentional engagements create for the actions and psychological resources of different individuals on different occasions, thus hopefully doing better service to our scientific and clinical practices. Much more needs to be done to further elaborate and defend the present account, but I hope the work done so far makes for a promising start.

Acknowledgements I am very grateful to Lucas Battich, Ophelia Deroy, Patricia Gera, Bart Geurts, Anpeng Liu, Will McNeill, Marat

Shardingaliev, and Harry Tappenden for their generous and perceptive comments, in conversation and on different versions of this work. I would also like to thank two anonymous reviewers for their especially thorough and helpful comments, which greatly improved the quality of the present work. Finally, I am grateful to the editors Anna Bloom-Christen and Michael Wilby for their kind assistance along the way.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The author has no conflict of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Battich L, Geurts B (2021) Joint attention and perceptual experience. *Synthese* 198:8809–8822
- Battich L, Fairhurst M, Deroy O (2020) Coordinating attention requires coordinated senses. *«Psychonomic Bull Review»* 27:1126–1138
- Behne T, Carpenter M, Tomasello M (2005) One-year-olds comprehend the communicative intentions in a hiding game. *«Developmental Science»* 8:492–499
- Bloom P (2002) *How infants learn the meanings of words*. MIT Press, Harvard
- Brandom RB (1994) *Making it explicit: reasoning, representing, and discursive commitment*. MIT Press, Harvard
- Bratman ME (1987) *Intention, plans, and practical reason*. Harvard University Press, Cambridge, MA
- Bruner JS (1974) From communication to language: a psychological perspective. *Cognition* 3(3):255–287
- Butterworth G (1995) Origins of mind in perception and anger. In: Moore C, Dunham P (eds) *Joint attention: its origins and role in development*. Routledge Psychology Press, New York, pp 29–40
- Campbell J (2005) Joint attention and common knowledge. In: Eilan N et al. (eds) *Joint attention: communication and other minds: issues in philosophy and psychology, consciousness & self-consciousness series*, pp 287–297
- Carpendale JIM, Carpendale AB (2010) The development of pointing: from personal directedness to interpersonal direction. *«Human Development»* 53:110–126
- Clark HH (1996) *Using language*. Cambridge University Press, Cambridge
- Csibra G (2010) Recognizing communicative intentions in infancy. *«Mind & Language»* 25:141–168
- Csibra G, Gergely G (2009) Natural pedagogy. *«Trends in Cognitive Sciences»* 13(4):148–153
- Drobňák M (2022) Normative inferentialism on linguistic understanding. *Mind & Language* 37(4):564–585

- Eilan N, Hoerl C, McCormack T, Roessler J (eds) (2005) Joint attention: communication and other minds. Oxford University Press, Oxford
- Gallagher S (2011) Interactive coordination in joint attention. In: A. Seemann (ed) Joint attention: new developments in psychology, philosophy of mind, and social neuroscience, Cambridge (MA): MIT Press. 293–306
- Geurts B (2018) Making sense of self-talk. «Review of Philosophy and Psychology» 9:271–285
- Geurts B (2019a) Communication as commitment sharing: Speech acts, implicatures, common ground. «Theoretical Linguistics» 45(1–2):1–30
- Geurts B (2019b) Commitments continued. «Theoretical Linguistics» 45(1–2):111–125
- Geurts B (manuscript) (2020) Common ground as a normative condition. Manuscript version: May 31
- Geurts B (under review), Common ground in pragmatics, manuscript version: Nov 19
- Geurts B (forthcoming), Normative pragmatics and social structure: an evolutionary perspective, in Evolutionary pragmatics, Oxford: Oxford University Press
- Griffin R, Dennett DC (2009) What does the study of autism tell us about the craft of folk psychology? In: Striano T, Reid V (eds) Social Cognition: Development, Neuroscience and Autism. Wiley Blackwell, Oxford, pp 254–280
- Hildebrandt F, Glauer R, Kachel G (2020) *Coming from a world without objects*, «Mind & Language»: 1–18
- Jones W, Klin A (2013) Attention to eyes is present but in decline in 2–6 month-olds later diagnosed with autism. *Nature* 504(7480):427–431
- Karpus J, Radzvilas M (2021) Game theory and rational reasoning. In: Heilmann C, Reiss J (eds) The routledge handbook of the philosophy of economics, New York: Routledge, 2021:99–112
- Kissine M (2021) Autism, constructionism, and nativism. *Language* 97(3):e139–e160
- Laidlaw KEW, Risko EF, Kingstone A (2016) Levels of complexity and the duality of gaze: how social attention changes from lab to life. In: Obhi SS, Cross ES (eds) Shared representations: Sensorimotor foundations of social life. Cambridge University Press, Cambridge, pp 195–215
- Lewis D (1969) Conventions. Harvard University Press, Cambridge
- Liszkowski U, Carpenter M, Henning A, Striano T, Tomasello M (2004) Twelve-month-olds point to share attention and interest. «Developmental Science» 7(3):297–307
- Ludwig K (2007) Collective intentional behavior from the standpoint of semantics. *Nous* 41(3):355–393
- Mole C (2011) Attention is cognitive unison: an essay in philosophical psychology, Oxford: Oxford University Press
- O'Madagain C, Kachel G, Strickland B (2019) The origin of pointing: evidence for the touch hypothesis. *Sci Adv* 5(7):eaav2558
- Peacocke C (2005) Joint attention: Its nature, reflexivity and relation to common knowledge, in Eilan 2005:298–324
- Reddy V (2008) How infants know minds. Harvard University Press, Cambridge (MA)
- Scarafone A (2023) What would it be like for prelinguistic communication to be Gricean? «Language and Communication» 90:1–13
- Scarafone A, Michael J (2022) Getting ready to share commitments. «Philosophical Topics» 50(1):135–160
- Schmidt MFH, Rakoczy H, Tomasello M (2019) Eighteen-month-old infants correct non-conforming actions by others. «Infancy» 24(4):613–635
- Stuart N, Whitehouse A, Palermo R, Bothe E, Badcock N (2023) Eye gaze in autism spectrum disorder: a review of neural evidence for the eye avoidance hypothesis. *J Autism Dev Disord* 53:1884–1905
- Tomasello M (2019) *Becoming human: a theory of ontogeny*. Harvard University Press, Cambridge (MA)
- Weiss B (2022) From tools to rules. «Philosophical Topics» 50(1):55–82
- Wilby M (2023) The form and function of joint attention within joint action. *Philos Psychol* 36(1):134–161
- Wu W (2014) *Attention*. Routledge, New York

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.