



Joint Attention: The PAIR Account

Michael Schmitz¹

Accepted: 28 February 2024
© The Author(s) 2024

Abstract

In this paper I outline the PAIR account of joint attention as a perceptual-practical, affectively charged intentional relation. I argue that to explain joint attention we need to leave the received understanding of propositions and propositional attitudes and the picture of content connected to it behind and embrace the notions of subject mode and position mode content. I also explore the relation between joint attention and communication.

Keywords Joint attention · Intentional content · Force/mode · Propositional attitudes · Nonconceptual content

1 Introduction

What is joint attention? As many have noted (León 2021; Harder 2022; Eilan [manuscript](#); Siposova and Carpenter 2019), while there has been a lot of valuable empirical psychological work on joint attention, particularly in developmental psychology, there is still significant conceptual unclarity about what it is. In this paper I will present an account of joint attention that I call the “PAIR account”: the account of joint attention as a perceptual-practical, affectively charged intentional relation. And I will also provide at least some indications why it has been so difficult to understand joint attention in the framework adopted by most mainstream philosophy and psychology—the framework of propositions and propositional attitudes.

Joint attention can be characterized as the union of two basic capacities. First, infants engage in exchanges of vocalizations, looks, smiles and other affectively charged mimicry with their caretakers or other people. These are often called “protoconversations” (see e.g. Trevarthen 2012). Second, they also explore the world of objects by playing with things and gazing at their surroundings. At around 9–12 months of age (Tomasello et al. 2005), or perhaps even earlier (see Moll 2023 for discussion), they begin to engage with objects together with others. They draw others’ attention to them (or have their attention drawn to them by others) to express and share their feelings about them.

On a common view in psychology, joint attention involves a pro-social motivation to share and typically has a tripartite structure of: (1) an initiating act to get the other’s attention, (2) a referential act of pointing to the object to be shared, (3) a “sharing look” or other communicative act which comments on the object by expressing a feeling about it like e.g. excitement, wonder, or concern (Carpenter and Liebal 2011). But joint attention can of course also be “bottom up”, as when the subjects’ attention is drawn by a stimulus they both experience like e.g. a loud noise.

Now contrast this with some of the early philosophical accounts of joint attention. These were inspired by accounts of common knowledge and appealed to a notion of the openness or epistemic transparency of a joint attention situation where two subjects are jointly aware of an object X. Stephen Schiffer characterized this perceptual openness as follows: “A perceives X, B perceives X, A knows that B perceives X, B knows that A perceives X, A knows that B knows that A perceives X, B knows that A knows that B perceives X”, and so on and so forth (Schiffer 1972, p. 30). But analyses that appeal to such infinite chains of epistemic, theoretical mental states seem questionable in several ways.

First, there is an obvious worry that they will lead to an infinite regress. One response to this is that there is no regress of actual mental states because the analysis just characterizes the mental states the co-attenders are disposed to form, so that the infinity is merely potential (Wilby 2010). However, it seems plausible that when co-attenders e.g. share a sharing look, they immediately experience the

✉ Michael Schmitz
Michael.Schmitz@univie.ac.at

¹ Universität Wien, Vienna, Austria

jointness, the closing of the triangle of joint attention,¹ and that their experiences of jointness are actual mental states that cannot be reduced to mere dispositions, or other forms of mere potentiality, as accounts that try to understand jointness in terms of mutual availability like Christopher Peacocke's (2005) appear to attempt.

Second, the analysis just seems overly complex and contrived (cf. León 2021; Eilan *manuscript*). The natural way for either participant to conceptualize and verbally express their experience—if they are able to do so, which importantly is not necessarily the case—would be to say something like “We know such and such to be interesting / funny” or “We see / hear such and such” (cf. Schmitz 2014, 2018 and Eilan *manuscript*), where *we* is the subject of the experience or the state of knowledge.

Third, because joint attention is commonly thought to emerge around 9–12 months, if not earlier, the fact that this kind of account requires recursive mindreading is deeply problematic, as it is controversial that at this age infants understand mental states at all, much less potentially infinite levels of recursion that even adults, including philosophers, psychologists and cognitive scientists, often find it difficult to wrap their minds around. Fourth, analyses of the kind under discussion try to understand joint attention solely in terms of epistemic, theoretical states like perception and theoretical knowledge or belief. But it's questionable that states of this kind by themselves can satisfactorily account for the jointness of joint attention.

Here's an example to make this point vivid: “Consider two people who are focused on the same target, a high-ranking politician. One wants to shoot him, the other, the politician's bodyguard, wants to protect him. The bodyguard tracks the assassin out of the corner of his eyes because he has become suspicious of her. The assassin also tracks the bodyguard's attention because if the bodyguard loses track of her, she will have the time to get her gun out and shoot the politician; otherwise the bodyguard could shoot her first.” (Schmitz 2014, p. 238; two obvious mistakes were corrected). So these two are attending to the same object and they are mutually aware of their attending up to whatever level we want to take it: the bodyguard knows that the assassin knows that he has perceived that she has perceived, and so on.

The example is also constructed to meet a causal requirement formulated by John Campbell to the effect that one's continued attention to the object “must be one of the factors

causally sustaining” (Campbell 2002, p. 162) the other's continued attention to it. Still, intuitively it does not seem correct to say that the assassin and the bodyguard are jointly attending to the politician. A similar—but less homicidal—example is given by Rory Harder (2022). He imagines a scenario where two people in a park covertly and mutually monitor their attention to a dog which is owned by one of them and a variation of this scenario, where the two express similar feelings of fondness for the dog through sharing looks, concluding that joint attention is present in the second scenario, but not in the first.

Let us distinguish three relevant concepts here—epistemic transparency or openness, mutuality and jointness—and clarify their relation.² I assume that in this context at least epistemic transparency is the same as openness, and that openness essentially contains a dispositional element. That something is out in the open means that it is *available* to the participants in some way. The contrast between mutuality and jointness is essentially the contrast between attitudes or relations subjects have towards each other and an attitude or relation they have together—meaning as a unit in some sense—toward a third thing. In our example, it would be true to say that the bodyguard and the assassin were mutually aware of each other as attending to the politician, but, the claim is, it would not be true to say that they jointly attended to the politician, because they did not form a unit in the relevant sense. Similarly, to jointly open the door would be something different from mutually opening the door for one another. Or, to use an example from the law, mutual wills are wills that contain reciprocal provisions concerning the respective parties, whereas a joint will is a single document that the parties draw up together, as a unit.

How are these concepts related? Openness as a dispositional notion can be taken to mean the availability of either mutual or joint awareness. The accounts of Schiffer and Peacocke are best understood as proposals to reduce joint attention to the open-ended, potentially infinite, mutual availability of awareness of attention. In concert with many recent positions in developmental psychology (e.g. Carpenter and Liebal 2011; Siposova and Carpenter 2019; Moll 2023) as well as in philosophy (e.g. Campbell 2002; Hutto 2012; Gallagher 2011; Schmid 2014; Seemann 2011) I believe that this reduction fails. Jointly attending is a non-dispositional, occurrent state of attending together as a unit. It's the joint experiential possession of an object such as a thing or state of affairs.

The crucial question of course is what ties the subjects together so that they form a unit. A central claim of this paper is that this tie cannot be merely epistemic or

¹ For a corresponding assessment of the deficiencies of the traditional kind of view, compare Siposova and Carpenter: “However, this level of attention and knowledge does not seem to be known truly together in any meaningful way. In contrast to the ‘meeting of minds’ that takes place in the classic ‘joint attentional triangle’, it is basically two individual perspectives that never meet in the middle: Each individual just assesses the attention and knowledge states of the other individually...” (2019, p. 262). See also Moll (2023).

² I thank an anonymous referee for the journal for the suggestion to clarify the relationship between these concepts.

theoretical. Merely being mutually aware of one's attention is not sufficient for such a tie and thus for attending jointly. If we are truly jointly aware, we will also act together or are at least disposed to act jointly. In most examples in the literature, we may engage in joint communicative actions such as the exchange of sharing looks. And these sharing looks also communicate a shared feeling about the object, in this case, fondness, and this feeling disposes the co-attenders to further joint actions such as petting the dog or playing with it. Moreover, the very urge to communicate is already expressive of a pro-social motivation. We will discuss later whether jointness is necessarily tied to communication.

First I want to address a widespread concern about the idea of attending as a unit and of joint subjects such as a *we*. Doesn't this mean that the unit must have a mind of its own, something like the dreaded 'group-mind'? But this is to misunderstand the nature of the connection. The *we* is essentially a *plural* subject. Jointness neither creates a separate third entity with a mind of its own, nor do the co-attenders fuse and so disappear into the new unit. (To put this into a slogan: no *we* without *Is*.) It's rather that the co-attenders mutually experience and represent themselves as being related to one another and to the objects they are attending to. As noted above, the most natural way for either of them to report their experience would be something like "We are watching this dog"—where with "we" the co-attenders represent each other as being related to one another in the special way we are trying to elucidate. In so doing, they also represent each other as co-subjects of this perceptual relation.

To avoid a misunderstanding, let me note that on the view to be developed, mutuality is still essential to jointness. If, for example, in the dog case one subject would turn away their gaze and refuse to participate in the sharing look, this wouldn't be an instance of joint attention even if the other subject still were in joint attention mode, that is, in the kind of mental state that, if both subjects were in it, would make this an episode of joint attention. What I am arguing is that mere mutual availability is not sufficient because we need occurrent representations, and that mutual representations of one's own and of others' attention *as objects* is not sufficient because the co-attenders need to mutually represent each other as *co-subjects* of the attention relation.

While I have appealed to examples where first-person plural pronouns like "we" are used in thought or speech, let me emphasize that with joint attention we are after a phenomenon that is more basic than any use of "we". What we are trying to understand are the basic, sensory-motor-emotional, forms of jointness. These can then form the basis for higher-level forms of collective intentionality which depend on these lower-level forms such as we-thought, we-speech and we-reasoning (cf. Tomasello (2014) as well as role-mode intentionality (Schmitz 2018; 2023), where

subjects act in social roles such as being a citizen or an employee.

Before we finally come to the task of characterizing jointness, we need to get another more negative or diagnostic task out of the way. The familiar and deeply entrenched understanding of intentional states and speech acts as "propositional attitudes" is a main source of skepticism about the phenomena of jointness and makes analyses along the lines we have discussed appear natural or even inevitable. In the next section I will characterize this framework. It is based on the dichotomy of force/mode and propositional content: force/mode as what makes the act e.g. an assertion or direction, or the state e.g. a perception, or intention, vs. content in the sense of *what* is asserted, perceived, or intended, which is taken to be a proposition and thus a truth-value bearer.

2 Joint Attention and the Received Framework of Propositions and Propositional Attitudes

This framework has many different aspects which I cannot all discuss or even list here.³ I focus on the following which I think make it particularly difficult to adequately understand joint attention.

- 1) The representational or intentional content of a posture—the propositional attitude or speech act—is generally taken to be identical to the embedded proposition. The subject and the attitude or speech act type, the mode or force, make no contribution to content. They can only become the object of a report, when the fact that a subject has such an attitude is represented as part of the propositional content of another posture.
- 2) When the subject of a posture is considered, it is usually taken for granted that this subject must be an individual, an I-subject. And this in turn is because it is thought that a collective or we-subject would have to be something like a group mind, and this idea is—rightly—considered to be preposterous.

With this, we can already give a diagnosis how the regress arises. Since according to (1) each co-attender and its attitude, in order to be represented, need to be represented as part of the propositional content of a report, the subject and force/mode of this report will again not be represented, will

³ But see Schmitz and Mras (2022). Historically, the framework mostly derives from Frege's (1956) notion of a thought. For a more recent formulation see Searle (1983), and for a helpful overview McGrath (2005).

be offstage as it were, so that we need another report, which has that subject and force/mode as part of its propositional content, which will generate another level of report, to capture which we have to move up yet higher in recursion, and so on, ad infinitum.

Moreover, even if we abandon (1), as I will argue we should, in favor of the idea that force/mode is itself representational, meaning that the subject is always aware of the position it takes up vis-à-vis the state of affairs (SOA) represented by what is traditionally known as propositional content, we still can't make sense of the joint epistemic possession of a SOA—a SOA that *we* jointly attend to or that *we* jointly know to be the case—if we can't use the idea of a collective subject which (2) rules out. Why should the proponent of the regress analysis be impressed by this? It seems to me that we should make sense of what seems intuitive if we can, especially given the steep theoretical costs of the recursive view already mentioned at the beginning.

I thus want to propose a view that in response to (1) and in a tradition that ranges from Kant to Piaget, Peter Strawson, Gareth Evans and contemporary philosophers like Jose Luis Bermudez, urges that there is an essential connection between world-consciousness and self-consciousness. In any posture ranging from perceptual and actional experience to the most abstract thought, a subject is never just aware of a SOA or other object, but always also of its own *position* vis-a-vis that object, where “position” can mean spatial position as in perception; but also causal position, as we experience the world acting on us in perception and us acting on the world in action; as well as temporal position, as we situate ourselves temporally, e.g. through tense; as well as our conative, cognitive or epistemic position, as, for example, we represent ourselves as knowing what is the case in assertion, or as occupying a certain practical position of being poised for action in intending. When this representation corresponds to the force/mode of the posture, I will refer to it as an instance of mode representation. Force/mode always represents the position from which a subject is aware of the relevant object. Note that this does not mean that the subject necessarily has a concept of this position. In fact in basic cases, position is rather represented nonconceptually as e.g. through intonation contour and word order in the case of assertion. But it does mean that in any posture the subject is also always aware of itself, that any posture has a moment of self-consciousness. That aspect of a posture I also call the “subject mode”. Subject modes include not only I-mode, but also we- and role-mode and, most importantly for present purposes, the mode of joint attention and action,

where the relevant self-consciousness is nonconceptual, as opposed to the conceptual self-consciousness manifest in the use of “I” and “we”.

The second essential claim I want to put forward—in response to (2)—is that there is a fundamental difference between representing others as objects of such positions and as co-subjects. I believe that co-subjective representation is the key to understanding jointness and collective intentionality more broadly.⁴ To experience or otherwise represent somebody as a co-subject is to experience or otherwise represent them *as being related to one in a certain way*. To understand collectivity in terms of co-subjectivity thus also means to reject the idea that collective subjects are somehow over and above the individual members of the collective. As we have noted already, that is not what jointness is about. It's rather that, for example, in exchanging a sharing look expressing their amusement or concern about a situation, both co-subjects will experience the other as sharing this response with them. They nonconceptually experience themselves as perceiving this situation and responding to it in a certain way, and the other as also responding to it in this way, thus closing the triangle of joint attention.

The third feature of the received doctrine of propositional attitudes that I want to address is its “theory bias”—its bias towards theoretical over practical forms of representation. The theory bias encompasses two distinct, but connected, biases. One is a bias for representations that are theoretical in the sense that they have mind-to-world direction of fit⁵ like perceptual experiences, beliefs and instances of theoretical knowledge, over representations that have world-to-mind direction of fit like actional experiences, intentions and instances of practical knowledge. This form of bias can also be called “cognitivist”. It is manifest in the privileged positions propositions have in the received view, as well as in many other popular doctrines such as truth-conditional semantics or the reduction of practical knowledge to theoretical knowledge of what is the case.

As truth value bearers, propositions belong to the theoretical domain, since truth is representational success from a theoretical position towards the world. Still, propositions are supposed to be the content of both theoretical attitudes like belief and practical attitudes like intention. I propose to rather think of the content that may be shared between an intention such as e.g. an intention to close the door and the belief that I will close it, as SOA content representing the SOA of me closing the door. But this same SOA can be represented as a fact, as something that is the case, from a theoretical position and as a goal, as something to do, from

⁴ Cf. Schmitz (2019). For a related distinction between the origin (subject) and target (object) of representation and a perspective on collective intentionality close in spirit to the present one, see also Shteynberg et al. (2023).

⁵ For the notion of direction of fit, see Anscombe (2000) and Searle (1983).

a practical position towards the world. SOA content itself is neutral between these positions, essentially incomplete and not truth-evaluable. To become truth-evaluable it needs to be supplemented by an indication that the SOA is represented from a theoretical position such as the indicative mood. So on the view I propose, any posture has SOA or object content, force/mode content that represents the subject's position and subject mode content because a subject cannot represent its own position without representing itself.

The second form of the theory bias is a bias for forms of representation that are propositional and conceptual over ones that are nonpropositional and nonconceptual. I use the term “intellectualism” to refer to it. The two biases are connected as on the traditional view, cognitive representation that is propositional and conceptual is the central and indeed, on many views, the only form of representation. And views that label themselves as intellectualist are typically also cognitivist. Often the cognitivist component is not even noted, but simply taken for granted. For example, in presenting their account of knowledge-how as a species of knowledge-that, Stanley and Williamson (2001) argue at length against Rylean accounts of knowledge-how as a nonpropositional, nonconceptual skill, but don't even consider the notion that there might be knowledge-how that has world-to-mind direction of fit and is expressed through imperative sentences and directive rather than assertoric speech acts, such as, for example, recipe knowledge of how to make Spaghetti Bolognese.

In accounts of joint attention, the cognitivist bias comes out in the tendency to think that joint attention should be entirely a matter of cognitive states like perception and belief (e.g. Schiffer 1972; Peacocke 2005) and the intellectualist bias in the tendency to think that their contents must be propositional and conceptual. I have already said why I think joint attention cannot be entirely explained in cognitive terms: cognitive states can only give us mutual awareness of attention as an object, but jointness must also involve a prosocial motivation and at least a disposition to also act jointly. To experience others as co-subjects, specifically as co-attenders, is thus what Ruth Millikan (1995) calls a pushmi-pullyu representation: a representation that has both theoretical, mind-to-world direction of fit and practical, world-to-mind direction of fit aspects.

The basic reason that intellectualism is inadequate is this: intellectual states are thought states, but it is implausible that joint attention is a matter of thought. It's rather a matter of experience: of perceptual experience—though at a level higher than, say, basic object perception—as well as of actional and emotional experience. Experience provides a more direct and immediate form of access to the world than thought, and the concept of nonconceptual content (e.g. Gunther (2003)) is a familiar tool for capturing what is special about these more basic forms of representation:

among other things, they are more context-dependent, independent of higher-level thought states such as belief and intention, and do not bring reflective abilities with them such that, for example, that a subject is able to attend jointly does not mean it is able to reflect on whether it is really attending jointly, or should be attending.

There are two more points I want to mention in connection with the received framework. They don't directly belong to it, but to the wider intellectual context of which it is a part. The wider context is that of an intellectual ideal that values propositional and conceptual representation of facts above everything else. This ideal is also, first, connected to an understanding of emotion where emotion is not only seen as itself not representational, but as something that impedes rather than enables adequate representation of the world: ideal representation is emotion-free. And second, ideal representation is also completely analyzed, that is, in terms of the ontologically basic constituents of the world.

These two notions have also been obstacles to an adequate understanding of joint attention. The first because of the importance of affect and emotion to an appropriate understanding of joint attention. At the basic level we are interested in here, the bond that ties human and other creatures together is emotional. This is common sense and borne out by joint attention research, some of which I will discuss later. The second point I want to make is connected to this because the level at which creatures relate to one another emotionally, perceptually and actionally as co-attenders is also one where in many ways their representations are not yet differentiated and analyzed, at least not in the way that conceptual level representations are. But this can be difficult to accept as we are very accustomed to thinking in terms of certain concepts, especially those that have a special significance for us, and often take for granted that it must be possible to specify all intentional contents in their terms. A case that is central to the present issue is that of the conceptual mind/body dualism. By the conceptual—as opposed to the metaphysical—mind/body dualism I here mean the notion that all representation must be as physical / bodily or as mental. This notion leads to the idea that in joint attention experience I must be experiencing either mere bodily behavior, from which mental states can only be inferred, or else I must directly perceive mental states. I believe that neither view is adequate and that in joint attention experience we experience others at a level prior to the mind/body distinction (Schmitz 2014).

3 Joint Attention Without Content?

Before I come to sketch the alternative PAIR account of joint attention I want to put forward, I want to briefly discuss some of the more recent accounts proposed in

the philosophical literature, namely the enactivist and particularly the relational account—both derived from corresponding accounts of individual perception. I agree with these accounts that joint attention is more basic than the so-called “propositional attitudes”. With enactivism I also agree insofar as it holds that there is an essential connection between joint attention and action; however, I do not think either individual or joint perception can be reduced to actions or dispositions to action (cf. Wilby 2023, p. 143f. for a corresponding critique of enactivism). And with John Campbell and others who adopt a “relational” and “naïve realist” view of the experience of individual and joint perception and attention, I agree that there are such experiential perceptual and attentional relations that hold between individual subjects, or jointly between them and their co-subjects and items in the world. I also agree with the naïve realist ideas that we experience the world directly and (mostly) as it is.

However, both enactivists and relationists are led by their views to reject the application of notions such as representation and content to individual and joint perception and attention. Campbell’s relationism has therefore been aptly called “austere relationism” (Schellenberg 2011). It is further strongly externalist in that it makes the existence of these experiences dependent on the existence of the corresponding objects as well as—for joint attention—of co-subjects with appropriate attitudes. It seems to me that there is a sense in which both actional and perceptual experience must be representational, and that it must be possible to decompose the experiential actional and perceptual intentional relations into the contributions made by individual subjects and their contentful states of consciousness and by the objects of these states, and that these individual mental states can and sometimes do exist without corresponding objects and co-subjects as in the bad cases of misrepresentation, of hallucination and illusion, including misrepresentations of jointness.

I will therefore try to articulate the sense in which perceptual experience must be representational and contentful. And I will argue as explicitly as I can⁶ that the existence of experiential intentional relations must depend on the existence of individual internal contentful states of experience. For example, co-subjects can only jointly experience a dog if each of them individually is in joint attention mode states with corresponding contents. But the

converse is not true: an individual may be in such a state while the other’s attention may have gone away or never have been there in the first place.⁷ Without disagreeing with this familiar argument against austere relationism from the bad cases, I do not want to merely repeat this argument here, but instead focus on the good cases. It is sometimes thought that these are unproblematic for austere views (e.g. Byrne and Green 2023), but it is important to show that this is not actually true.

One more clarificatory remark: often a distinction between presentation and representation is made, such that e.g. perception is presentational, while conceptual and propositional states such as belief are representational—as they may at least re-present a SOA that has been present to the subject before. I’m happy to accept a distinction along these lines, but (following Searle 1983) terminologically I prefer to use “representational” as a cover term for presentational as well as re-presentational forms of intentionality. I also believe that this distinction as well as related distinctions like between being acquainted with something in the world vs. merely knowing it by description, should be explained in terms of differences in content so that, for example, a presentational state or one of being acquainted involves nonconceptual content, but one of belief or propositional knowledge conceptual content.

The reasoning why attentional intentional relations require content can be spelled out as follows, using the central example of perceptual relations:

1. What a subject perceives does not only depend on its environment, but also on how this environment affects the subject’s organism. For example, whether a subject can see the letters on a screen or hear a high-pitched sound depends on its visual and auditory acuity and thus on how sights and sounds affect its nervous system, especially the visual and auditory areas of its brain.
2. When the relevant perceptual relation is also an experiential relation, the effects of the visual and the auditory stimulus must include corresponding visual or auditory experiential states of the organism, in our example conscious correlates of (part of) the activity in the visual and auditory brain areas. This is because the subject cannot be said to *experience* the sights or sounds without being in corresponding experiential states. Like their neuronal correlates, these states are internal states of the organism.⁸
3. The visual and auditory experiential states fundamentally differ regarding which features of the world they let us experience / reveal to us / acquaint us with / represent.

⁶ Thanks to an anonymous referee for this journal for pushing me to make my argument more explicit here.

⁷ The obvious point that a subject may be in “the psychological state of shared attention” without an actual co-attender is also made by Shteynberg (2015, p. 581). For critical discussion of Campbell’s view see also Cochrane (2009, p. 63f).

⁸ See Sinhababu (2023) for an argument that experiential features must be co-located with their neuronal correlates.

The auditory experience presents auditory features, the visual experience visual ones. And the experience of a high-pitched tone is also correspondingly different from that of a low-pitched tone, that of red from that of green, that of a co-subject from a mere object, and so on. Now that feature of an experience that relates us to / puts us into contact with certain features of reality, but not others—if it so puts us into contact and not merely seems to do so—is called “content”. (For purposes of this argument, this can be considered a stipulative definition.)

The argument proceeds from (1) that intentional relations must depend on inner states of the subject, to (2) that experiential intentional relations must depend on inner experiential states, to (3) that different inner experiential states relate us to different objects and thus differ in content.

Naïve realists such as Campbell and Fish seem to implicitly accept (1) because they emphasize that either physiological (Fish 2009) or subpersonal “cognitive processing” states (Campbell 2002, p. 118) are enabling conditions for experiential relations. (And I think anybody would find (1) hard to deny.) They would likely attempt to block the inference from (1) to (2) by rejecting the distinction between experiential relations and (inner) experiential states.⁹ But a merely physiological or information-processing state cannot underwrite an experiential relation to the world: that requires an experiential state. Only if the state that puts the subject in contact with the object is experiential, is a state of consciousness, can the subject be said to experience this object or to be conscious of it. And if the relationist rejects any notion of experience as an internal state, this is not only extremely implausible, but leaves mysterious both what experience really is and where it is.

Is it literally external to the mind as Campbell’s well-known remark that the “phenomenal character” of experience is “constituted by the layout and characteristics of... external objects” (ibid., p. 116) seems to suggest? If we go with this interpretation, it appears that the phenomenal character of experience is—quite surprisingly—mind-independent. Or else, if it is not, the question arises which feature of the subject’s mind it is dependent on. Again it would seem that only an inner experiential state could qualify. But if the reality of inner experiential states is accepted after all, we can simply ascribe phenomenal character to them rather than to anything external. Moreover, if inner experiential states are accepted as real, it is also hard to deny (3), that is, that they differ in ways that determine which environmental features they reveal and relate their subjects to—and so are contentful by definition.

⁹ This response was also suggested to me by an anonymous referee for the journal.

If this is on the right track, experiential relations to the world require contentful inner experiential states. By the same light, it is also not possible to be a naïve realist about the world part of the experiential relation alone—as the austere view seems to try to. One must rather be a naïve realist about consciousness and intentionality, too, and embrace the view that subjects experience features of the world because contentful inner states of consciousness put them into contact with them. Such a thoroughgoing naïve realism is adopted here.

But why are some theorists even tempted by austerity, why do they even try to be naïve realists about the world part of the experiential relation only? It’s not obvious that there is any conflict at all between naïve realism construed as the view that we experience the world directly and (mostly) as it is and the idea that there are inner contentful experiential states.¹⁰ A diagnosis of why some feel they are in tension would be desirable to complete the argument. For lack of space, I can only make some brief remarks here.

One set of reasons has to do with how the notion of content in some people’s minds remains connected to that of a proposition and even of being a representation in the sense of having a formal syntactic structure. Here I will just simply say that we must leave behind notions of content and representation that are biased towards higher-level propositional and linguistic representation and adopt a more inclusive concept which allows us to develop a unified framework for different forms of representation. The second, and even more influential, set of reasons or worries is epistemological. The first and crudest worry is that content will turn out to be a mental object rather ironically blocking direct access to the world rather than enabling it. (This worry is still connected to the notion of a propositional attitude, because if perception is a propositional attitude, then on many interpretations this means that the proposition is the object of the perceptual relation.) The second worry is that content can be known independently of any reference to or knowledge of the external world.

Consider the following telling analogy Campbell uses in his argument against representationalism: “It would plainly be a mistake to hold a Representationalist View of panes of glass: to hold that the only way in which it can happen that you see a dagger through a pane of glass is by having a representation of the dagger appear on the glass itself.” (2002, p. 118). The metaphor of the glass and its transparency is revealing because the experiential intentional state is cast in the role of an object—a glass with a (presumably pictorial) representation on it—so that it would indeed stand

¹⁰ This point is also made by Byrne and Green (2023), Logue (2014) and McDowell (2013). Schellenberg (2011) and Siegel (2010) are also “compatibilists” regarding content and relationism/naïve realism.

between subject and object and would block direct access to the world. The representation of the dagger on glass would have to be a visual object that can be seen. Such an account of visual experience is surely a mistake, but this is a point that defenders of intentionalism / representationalism about perception have also often made (e.g. Byrne and Green 2023; Logue 2014; McDowell 2013; Searle 1983; 2015). A visual experience is not even the kind of thing that can be seen; it is rather what enables the seeing of anything. It is not a visual object, but a subjective state which puts the subject in touch with visual objects.

Representationalists thus explicitly reject the view of content as a mental object, and no reason has been given why they would be committed to it. I suspect that Campbell and others have trouble of making sense of content as a property of internal, experiential intentional states, because they conceive of it on the model of a private inner object “whose intrinsic character is independent of the environment” (Campbell 2002, p. 119), and that therefore could be known or apprehended independently of any knowledge of the external world, and where awareness of this inner object would have to be the basis for such knowledge. But again, insisting on a distinction between content and object does not commit the intentionalist to the claim that content can be known independently of any reference to or knowledge of the external world. In the words of Wittgenstein (2003), when we explain the meaning of “red” even as it occurs in a statement like “This is *not* red”, we still point to something red (PI, § 429). He could have added that this point also applies to statements ascribing illusions or hallucinations involving redness—or any other feature—or indeed any statements merely reporting the contents of experiences, statements such as that it seemed or appeared to a subject that something was red.

On a tempting way of thinking, the choice in the philosophy of perception is between a view that says perception is fundamentally a relation and one that says perception is fundamentally a state present in the good—relational—cases, as well as in the bad ones, where the state fails to adequately relate its subject to the world. I have argued that we cannot make sense of an experiential relation without inner experiential states with contents that are distinct from the object the subject is related to, but are precisely what puts the subject in experiential contact with that object. I think the converse argument that we cannot make sense of such contentful inner experiential states without also ascribing experiential relations to the world to us can also be made along the lines suggested by the Wittgenstein quote.

To work this out in detail is beyond the scope of this paper, but here is, in a nutshell, one way of taking this further: at the level of perceptual experience, subjects represent themselves in relation to their environment, but in a way that

is prior to the mind/body differentiation and the distinction between object and content. That distinction is only acquired by understanding misrepresentation. But one can only self-ascribe a misrepresentation like an illusion by simultaneously taking oneself to also correctly represent the world. For example, in self-ascribing the Müller-Lyer illusion one takes oneself to know that the lines are equally long, and that knowledge is also based on perception. So the bad cases of misrepresentation can only be understood as deviations from the normal case of successful representation, and so (merely) contentful states presuppose perceptual relations to the world (see Schmitz 2019).

4 Joint Attention and Communication

Let us return to the question what is missing in the counterexamples to theory-biased, purely perceptual and/or propositional accounts of joint attention discussed above. So far I have only appealed to the notion of pro-social motivation and gestured towards a practical component, a disposition for joint action. One answer that has recently been gaining in popularity is communication (Eilan *manuscript*, Harder 2022, Moll 2023).

Is communication necessary for joint attention? Communication does seem very central to the paradigm cases of joint attention such as they are found particularly in the literature in developmental psychology—cases where an infant initiates a joint attention episode by a pointing or similar gesture and the episode concludes with a sharing look. It seems plain or at least plausible that communication is essential to these kinds of cases. However, such cases may be special in that they appear to involve not only jointness, but a particular act of directedness at an object for its own sake, which is removed from the flow of interaction, seems almost aesthetic in character and may indeed be best described as “proto-aesthetic”.

If we consider different kinds of examples, the thesis that joint attention must involve communication becomes much less clear-cut. Think of cases where an infant or adult initiates a joint action or joins that action simply by starting to act. For example, I may kick a ball to you, you kick it back and we start kicking back and forth. In so doing, I take it, we will also be jointly attending to the ball. It seems to me to be compelling to characterize our attention as joint because we will not only be mutually aware of the other as looking at the ball, but this looking will also be in the service of our joint goal of keeping it in play.

Such episodes typically do involve clearly communicative acts such as inviting looks or gestures or comments on the performance, expressions of joy or disappointment, and so on. But are they necessary? I think sometimes such episodes occur without separate communicative acts. Now, one

might respond by saying that kicking a ball to somebody in a playful and at least somewhat friendly manner—because surely not all ways of kicking a ball to somebody will be meant or perceived as initiations of joint play—is already communicative. I think there is a sense in which this is true, but that such cases are still importantly different from those that involve outright, clearly separable, communicative acts such as pointing gestures, which are typically appealed to in the literature.

Is communication sufficient for joint attention—when added to the counterexamples or cases of the same kind? That communication is not sufficient can be brought out by considering cases of antagonism and disagreement. If the assassin from the example above were to shout something derogatory about the politician to the bodyguard, this would surely not be sufficient to turn this episode into one of joint attention. Similarly, if you engage somebody to share a look of appreciation or amazement at something only to find that this person is concerned or even horrified about what you have perceived, this will at least not be a paradigmatic instance of joint attention (similarly Eilan, [manuscript](#)). Joint attention just like jointness in general thrives on like-me intentionality (cf. Meltzoff 2007): on imitation, attunement and agreement. Occasional disagreement is fine and may even strengthen the bond by adding a bit of spice or frisson to the relationship. (Interestingly, sometimes optimal ratios between positive and not so positive interactions have been proposed for romantic relationships. I suspect something like this is true for other kinds of relationships as well.) But if they disagree too often, the subjects will disengage, and their bond will be compromised or dissolve entirely.

I want to propose that the emotional bond, the connection, or communion (similarly Eilan [[manuscript](#)], who uses “commune”) created through agreement in theoretical and practical interests and proclivities of creatures is what ties them together as co-subjects and sustains episodes of joint attention even in the absence of outright communicative acts, as when, for example, we are looking at a sunset together, are watching a movie together, or are jointly listening to music. In this way we can also explain joint aesthetic appreciation, even though aesthetic appreciation is often thought to be characterized through a detachment from the world one might think is incompatible with the pro-social motivation essential to jointness on the present view.¹¹

Now, if the subjects are not in any way interested in sharing their experiences or having their experience shaped

through others, or shaping theirs, through joint engagement with the music or other aesthetic object, then indeed they are not really interested in joint aesthetic appreciation—though they might still enjoy going to the opera or museum together for the sake of having company on the walk there, or sharing a drink or meal later, etc.. However, this hardly always the case. An elementary and cross-culturally pervasive way of jointly engaging with music and mutually shaping one’s experience of it is dancing together or other forms of movement to music such as snapping one’s fingers, tapping the beat, and so on. Through the way a subject moves, it may reveal new features of the music to the co-subjects it is listening with: “dancing reveals aesthetic understanding” (Zangwill 2012, p. 388). In turn, it may also pick up new things from the others and discovering the music together and jointly responding to it will tie the subjects together: they will bond over their shared experience, or deepen an already existing bond. (Conversely, as noted already, failures to coordinate and synchronize perceptions and movements may weaken an existing bond.)

It is crucial here that the pro-social motivation consists in enjoying the enhancement of one’s aesthetic experience through jointness. Note that it is also not required that the subjects be reflectively aware of the causes of their enjoyment of the experience, or deliberately aim to have their listening experience shaped through joint engagement (pace Zangwill 2012, p. 386f), which would require having corresponding concepts. They may enjoy having their experience shaped through joint engagement with others without being able to conceptualize this and deliberately, reflectively aiming for it. The present view can thus make sense of joint aesthetic appreciation and the pro-social, emotional motivation inherent in it. There may be an ideal that aesthetic experience *should* be solitary that is itself aesthetic and part of the already mentioned individualistic tradition which tends to view others and emotional bonds with others as essentially an obstacle to both adequate cognition and adequate aesthetic appreciation. But this is not how people always feel about aesthetic experience—including, I suspect, adherents of that tradition.

Moving to music is another good example for an act that is not necessarily outright communicative like a speech act or pointing gesture, but that has a communicative aspect and, when done with others, can be a form of communion. But the jointness is not only manifest in such acts, it can also be manifest in how we experience the world even in the absence of such interactions, in how we look at the world with eyes that our sensitive to the needs, desires and fears of our co-subjects. For example, when I run with my wife, I’m much more sensitive to the presence of dogs, perceive them differently and am much more disposed to avoid them, because my wife is afraid of them. And I’m much more likely to notice things or events that we have interacted with

¹¹ Thanks to an anonymous reviewer for the journal for raising this concern and pointing me to a paper by Nick Zangwill (2012), which specifically argues against the notion of listening together. Zangwill responds to Cochrane (2009), whose account is more congenial to the present one.

in the past or that I sense she might find amusing, interesting or moving. This is how as co-subjects we look at the world with each other's eyes, open up new aspects of the world to one another and extend our theoretical, perceptual and practical, actional reach in it.

Is the difference between representing somebody as a co-subject and as a mere object a difference in content?¹² I think not primarily, because primarily the difference between subject mode and SOA content representation is structural. To jointly attend to the world is to attend to it from a position of identification with one's co-subject. And this is reflected in the structural role the co-subject has in revealing the world to me and in how I also view the world with their eyes in joint attention. At the same time, this structural difference is unthinkable without a difference in content. That is, if I do not experience you as revealing the world to me and as trustworthy regarding my interests etc., there must also be differences in content in comparison to somebody I do so experience.

5 The PAIR Account of Joint Attention¹³

I have been working towards an account of joint attention that I now want to state concisely, using the "PAIR" abbreviation as a mnemonic device to make it easier to remember. I will also present some empirical results both to support the account and to make more concrete what its key notions mean. In a nutshell, the PAIR account conceives of joint attention as a perceptual and practical or pragmatic and affectively charged intentional relation. That is, as against the theory-biased traditional accounts that view joint attention as a mere cognitive affair, as a matter of mere perception or belief and as propositional and conceptual, the PAIR account urges that joint attention cannot be a matter of mere mutual perception or awareness but must involve a pro-social motivation and corresponding dispositions. We also discussed the suggestion that communication is the central ingredient that turns mere mutual awareness into full-blown joint attention. While communication is certainly very important and may be necessary for joint attention if construed broadly enough to include what we have called "communion", we also found reasons to doubt that it is sufficient. Mere communication is not enough, but joint attention also requires agreement and affirmation among the co-attenders and generally what may be referred to as "like-me intentionality". And the jointness of joint attention can also be manifest in the absence of communication and even in the absence of mutual perception, namely in

how the co-attenders experience the world in a way that is sensitive to the interests, needs and feelings of the other. The identification with the interests, needs and feelings of the other is also what provides the affective charge in the communication and interaction with the co-attender and is part of what relates us to our co-attenders as co-subjects rather than as mere objects.

As against austere relationalist and radically enactivist accounts of joint attention that reject any appeal to the notion of intentional content, I have argued that such appeal is necessary to make sense of joint attention as an intentional experiential relation that relates the co-attenders to one another and to the objects in the world that they jointly attend to. An intentional experiential relation can only obtain in virtue of the content of the co-subjects' experience. Their merely physiological or 'sub-personal' information-processing states are not sufficient to determine an experiential intentional relation. But the contents of joint attention experience are nonconceptual as the contents of perceptual, actional and emotional experience are generally.

Let me conclude by discussing some findings from the literature in developmental psychology that support the general approach that I have sketched in the sense that they can be easily motivated and explained from the point of view of that approach. Many insights into how others are experienced, understood and treated in episodes of joint attention comes from research into the differences between autistic and neurotypical children. For example, when asked where a sticker should go, all non-autistic children in a study by Hobson and Meyer (2005) indicated that place by pointing to their own bodies, while more than half of the autistic children never indicated the place in this way, but always pointed to the place on the other's body. These different ways of pointing exemplify the difference between a co-subjective and an objectifying style of reference. To point to a place on one's own body to pick out the corresponding place on that of the other is to treat them as somebody "like me" rather than as an object.

Peter and Jessica Hobson have also found that there is a correlation between the frequency of sharing looks and role reversals in joint action, concluding that "the results suggest that the mode of social perception that involves sharing looks [also] gives rise to self-other transpositions in imitation (2011, p. 124). The PAIR account can explain this as a consequence of experiencing the other as a co-subject, as somebody who is like me, because people who are like me can perform the actions that I perform, and because I experience myself as forming a subject of action together with the other, so that it does not matter so much who does what, and we can easily switch between different roles in the pursuit of our shared goals.

Autistic children also engage much less in the kind of affirmative nodding people often engage in when listening

¹² I thank an anonymous referee for the journal for asking this very helpful question.

¹³ This section draws on material in Schmitz (2014)

to others. And only 3 of 16 children with autism showed a concerned look when a drawing by a tester, who they were in a joint attention relation with—or at least something that would be a joint attention relation for neurotypical children—was torn in their presence, while almost all neurotypical children did express concern for the tester (Hobson et al. 2009). This shows that autism is also connected to deficits in affirming the positions of others, and in experiencing the world regarding the feelings, interests and concerns of others. These results support the theses of a deep connection between joint attention and bonding through sameness and identification, and of a deep connection between subject- and object awareness.

That is, joint attention means that the co-subjects are attuned regarding cognitive or theoretical and conative or practical interests, as well as aligned regarding their physical features and stances as in mimicry. It is also manifest in how we often experience the world in relation to us and our common ground of shared interests and past experiences. Another result from developmental psychology nicely illustrates and supports this point. Infants shared several toy ducks with one experimenter and then several teddy bears with another. When they then entered a room with just one of the experimenters, in which a duck and a teddy bear picture were on the wall, they were much more likely to point to the picture of the object they had earlier shared with the experimenter they were with (Liebal et al. 2009).

There is some evidence that subject mode rather than SOA content explains certain kinds of social understanding and certain social actions based on that understanding. For example, 14-months-old infants understood an ambiguous request by an adult based on a shared joint attention episode, but not by merely observing his otherwise identical interactions with the relevant objects. After the adult and the infant had shared two objects and the infant had explored one object alone, the infant was able to correctly interpret an ambiguous request for “that one”, made with an excited expression by the adult, as referring to the new object. But 14-months-old infants were not able to do the same in conditions where infants merely observed e.g. the adult examine the objects by himself, or the adult engaging in joint attention with another person (Moll et al. 2007). Moll and Meltzoff conclude that “joint engagement is thus at least helpful, if not necessary, for infants of fourteen months to register others as becoming familiar with something” (Moll and Meltzoff 2011, p. 397).

It seems to me therefore that there is at least some *prima facie* support for the hypothesis that there is an important and multi-faceted difference between experiencing others as objects and experiencing them as co-subjects and for

the PAIR account of joint attention of which this hypothesis forms a core part.

Acknowledgments I thank the editors Anna Bloom-Christen and Michael Wilby for their work in putting this collection together, two anonymous referees for the journal for their very helpful comments and my wife Zuzana Toth for her love and support.

Funding Open access funding provided by University of Vienna.

Declarations

Conflict of interest The author declares that they have no competing interests.

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

- Anscombe GEM (2000) *Intention*. Harvard University Press, Cambridge
- Byrne A, Green EJ (2023) Whither naive realism?—I. *Philosophical Perspectives*. Wiley, Hoboken, pp 1–20
- Campbell J (2002) *Reference and consciousness*. Clarendon Press, Oxford
- Carpenter M, Liebal K (2011) Joint attention, communication, and knowing together in infancy. In: Seemann A (ed) *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience*. MIT Press, Cambridge, pp 159–182
- Cochrane T (2009) Joint attention to music. *Br J Aesthet* 49(1):59–73
- Eilan N (manuscript) Joint attention and the second person. Draft available online: <https://warwick.ac.uk/fac/soc/philosophy/people/eilan/jasphum.pdf>
- Fish W (2009) *Perception, hallucination, and illusion*. Oxford University Press, Oxford
- Frege G (1956) The thought: a logical inquiry. *Mind* 65(259):289–311
- Gallagher S (2011) Interactive coordination in joint attention. In: Seemann A (ed) *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience*. MIT Press, Cambridge, pp 293–304
- Gunther YH (2003) *Essays on nonconceptual content*. MIT Press, Cambridge
- Harder R (2022) Joint attention and communication. *Inquiry*. <https://doi.org/10.1080/0020174X.2022.2074101>
- Hobson RP, Hobson J (2011) Joint attention or joint engagement? Insights from autism. In: Seemann A (ed) *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience*. MIT Press, Cambridge, pp 115–136
- Hobson RP, Meyer JA (2005) Foundations for self and other: a study in autism. *Dev Sci* 8(6):481–491

- Hobson J, Harris R, García-Pérez R, Peter Hobson R (2009) Anticipatory concern: a study in autism. *Dev Sci* 12(2):249–263
- Hutto DD (2012) Elementary mind minding, enactivist-style. In: Seemann A (ed) *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience*. The MIT Press, Cambridge. <https://doi.org/10.7551/mitpress/8841.003.0016>
- León F (2021) Joint attention without recursive mindreading: on the role of second-person engagement. *Philos Psychol* 34(4):550–580
- Liebal K, Behne T, Carpenter M, Tomasello M (2009) Infants use shared experience to interpret pointing gestures. *Dev Sci* 12(2):264–271
- Logue H (2014) Experiential content and naive realism: a reconciliation. In: Brogaard B (ed) *Does perception have content?* Oxford University Press, Oxford
- McDowell J (2013) Perceptual experience: both relational and contentful. *Eur J Philos* 21(1):144–157
- McGrath M, Devin F (2005) Propositions. In: EN Zalta, U Nodelman (eds) *The stanford encyclopedia of philosophy* (Winter 2023 Edition). <https://plato.stanford.edu/entries/propositions/>.
- Meltzoff AN (2007) Like me: a foundation for social cognition. *Dev Sci* 10(1):126–134
- Millikan RG (1995) Pushmi-pullyu representations. *Philos Perspect* 9:185–200
- Moll H (2023) What we do and don't know about joint attention. *Topoi* October. <https://doi.org/10.1007/s11245-023-09961-y>
- Moll H, Meltzoff AN (2011) Joint attention as the fundamental basis of understanding perspectives. In: Seemann A (ed) *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience*. MIT Press, Cambridge, pp 392–412
- Moll H, Carpenter M, Tomasello M (2007) Fourteen-month-olds know what others experience only in joint engagement. *Dev Sci* 10(6):826–835. <https://doi.org/10.1111/j.1467-7687.2007.00615.x>
- Peacocke C (2005) Joint attention: its nature, reflexivity, and relation to common knowledge. In: Eilan NM, Hoerl C, McCormack T, Roessler J (eds) *Joint attention: communication and other minds*. Oxford University Press, Oxford, pp 298–324
- Schellenberg S (2011) Perceptual content defended. *Noûs* 45(4):714–750
- Schiffer SR (1972) *Meaning*. Clarendon Press, Oxford
- Schmid HB (2014) Plural self-awareness. *Phenomenol Cogn Sci* 13(1):7–24
- Schmitz M (2014) Joint attention and understanding others. *Synth Phil* 29(2):235–251
- Schmitz M (2018) Co-Subjective consciousness constitutes collectives. *J Soc Phil* 49(1):137–160
- Schmitz M (2019) The good, the bad and the naive. In C Limbeck-Lilienau, F Stadler (eds) *The philosophy of perception*. Berlin, De Gruyter, pp 57–74
- Schmitz M, Mras G (eds) (2022) *Introduction*. In: *Force, Content and the Unity of the Proposition*. Routledge, New York, pp 1–14
- Schmitz M (2023) From we-mode to role-mode. In M Garcia-Godinez, R Mellin (eds) *Tuomela on sociality*. Palgrave Macmillan, Cham, pp 177–200
- Searle JR (1983) *Intentionality: an essay in the philosophy of mind*. Cambridge University Press, Cambridge
- Searle JR (2015) *Seeing things as they are: a theory of perception*. Oxford University Press, Oxford
- Seemann A (2011) Joint attention: toward a relational account. In: Seemann A (ed) *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience*. MIT Press, Cambridge, pp 183–202
- Shteynberg G (2015) Shared attention. *Perspect Psychol Sci* 10(5):579–590. <https://doi.org/10.1177/1745691615589104>
- Shteynberg G, Hirsh JB, Wolf W, Bargh JA, Boothby EJ, Colman AM, Echterhoff G, Rossignac-Milon M (2023) Theory of collective mind. *Trends Cogn Sci* 27(11):1019–1031
- Siegel S (2010) Do visual experiences have contents? In: Nanay B (ed) *Perceiving the world*. Oxford University Press, Oxford
- Sinhababu N (2023) Qualia share their correlates locations. *Synthese* 202(2):1–14
- Siposova B, Carpenter M (2019) A new look at joint attention and common knowledge. *Cognition* 189:274. <https://doi.org/10.1016/j.cognition.2019.03.019>
- Stanley J, Williamson T (2001) Knowing how. *J Philos* 98(8):411–444
- Tomasello M (2014) *A natural history of human thinking*. Harvard University Press, Cambridge
- Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behav Brain Sci* 28(5):675–691
- Trevarthen C (2012) The generation of human meaning: how shared experience grows in infancy. In: Seemann A (ed) *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience*. The MIT Press, Cambridge. <https://doi.org/10.7551/mitpress/8841.003.0006>
- Wilby M (2010) The simplicity of mutual knowledge. *Philos Explor* 13(2):83–100
- Wilby M (2023) The form and function of joint attention within joint action. *Philos Psychol* 36(1):134–161. <https://doi.org/10.1080/09515089.2022.2039384>
- Wittgenstein L (2003) *Philosophical investigations: the German text, with a revised english translation*. Blackwell, Malden
- Zangwill N (2012) Listening to music together. *Br J Aesthet* 52(4):379–389

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