

Michael Schmitz

University of Vienna, Institute for Philosophy, Universitätsstraße 7, AT-1010 Vienna
michael.schmitz@univie.ac.at

Joint Attention and Understanding Others*

Abstract

In this paper I criticize theory-biased and overly individualist approaches to understanding others and introduce the PAIR account of joint attention as a pragmatic, affectively charged intentional relation. I argue that this relation obtains in virtue of intentional contents in the minds of the co-attenders, and – against the received understanding of intentional states as propositional attitudes – that we should recognize what I call ‘subject mode’ and ‘position mode’ intentional content. Based on findings from developmental psychology, I propose that this subject mode content represents the co-attenders as co-subjects, who are like them and who are at least disposed to act jointly with them. I conclude by arguing that in joint attention we experience and understand affective, actional and perceptual relations at a non-conceptual level prior to the differentiation of mind and body.

Keywords

joint attention, collective intentionality, other minds, intentionalism, mind-body dualism

1. Understanding others and the theory bias

In recent decades, mainstream philosophy of mind and psychology in general and the theory of understanding others, of social cognition, in particular have tended to be under the influence of three kinds of bias. Two of these can be brought under the heading ‘theory bias’, where, however, talk of the theoretical is taken in two different meanings. First, ‘theoretical’ can mean mind-to-world direction of fit speech acts and intentional states – I will from now on use ‘posture’ as a cover term for both – like assertions, beliefs and perceptual states, as opposed to world-to-mind direction of fit postures like requests, intentions and actional states. Examples for the theory bias in this sense include, but are by no means limited to, the idea that truth-conditional semantics can serve as a general theory of meaning, belief accounts of intention, and various forms of cognitivism in metaethics. Second, ‘theoretical’ can also mean

*

I would like to thank audiences in Berkeley, Helsinki and Vienna for helpful comments on earlier presentations of parts of this material, in particular Katharina Bernhard, Jennifer Hudin, Aloisia Moser, Hans Bernhard Schmid, John Searle, Gerhard Thonhauser, and Raimo Tuomela. Further thanks to Christian Neitzel and Josef Ehrenmüller for written comments on the penultimate draft and to Zdravko Radman for being a patient, cheerful, and sup-

portive editor. I would also like to thank the participants of a session at academia.edu on an earlier draft, in particular Bernard Baars, Jonny Blamey, Olle Blomberg, Anna Moltehanova, Nikos Psarros and Guido Seddone. Last, but not least, I thank the two reviewers for *Synthesis philosophica* for very detailed and helpful comments, and Kirsten Donaghey for improving my English.

– and in common parlance that is its prevalent meaning – something that is intellectual and discursive, that is propositional and has conceptual structure. The theory bias in this sense has often taken the form of the idea that intentional capacities of all kinds, ranging from actional and perceptual capacities to those for understanding language and other people, are to be explained in terms of an underlying theory – usually thought to be unconscious or even nonconscious. This intellectualist notion has been extremely influential in AI, cognitive science, psychology and philosophy, again in the idea of a truth theory as a theory of meaning (Davidson 1984), and of course in the theory-theory of social cognition, among many other examples. The third bias I have in mind is the bias for individual over collective forms of intentionality. This bias we can simply see reflected in the fact that until fairly recently, there was not much theorizing about collective intentionality at all and that much of that theorizing still tries to reduce the “we” to the “I”.

If we were looking for a single embodiment of these biases, we could find it in the Quine-Davidson-Dennett tradition of thinking about intentionality and language in terms of the thought experiment of radical translation or interpretation, where a linguist visits a community (a “tribe”) and translates their language from scratch. The understanding she gains is thought to be theoretical in both senses of the term: it consists in a translation manual with statements about what the words of the interpreted language mean. Moreover, the point of ascribing meanings and mental states is also thought to be theoretical – it is only for the purpose of explaining and predicting the behavior of those we interpret. Finally, meanings and mental states are ascribed reluctantly – if at all – because it is thought that while behavior is publicly accessible and thus epistemically respectable, mental states can at best be reasonable theoretical postulates – in pretty much the same the way that physicists are often thought to postulate ‘theoretical entities’ in order to explain observable going-ons. And the preferred option is to interpret such ascriptions in an instrumentalist or fictionalist way.

It is certainly no accident that in this picture the interpreter is an outsider and a scientist, somebody who comes into a community as a stranger and observes it from a detached, theoretical point of view. In fairness to this tradition it should be noted though that at least Davidson in his later work (Davidson 2001) moves more towards a view which emphasizes such points as that in order to share a language with people one also needs to share many beliefs and values with them and to relate to them in second-personal rather than third-personal sorts of ways.

In this paper, I will approach the topic of understanding others from a diametrically opposed starting point, namely that of joint attention. In joint attention we do not take a detached perspective on others, but rather attend to something with them. And our understanding of our co-attenders does not take the form of beliefs about what they perceive, think or mean, or of other propositional attitudes about them. Rather, it is manifest in actions and particularly in joint, co-operative actions. And it is manifest in how we experience our co-attenders and relate to them in the first place. So this understanding is not purely theoretical in the first sense, because it is manifest in both mind-to-world and world-to-mind postures, and it is not theoretical in the second sense, because all relevant sensory-motor-emotional experiences have non-conceptual, non-propositional intentional contents. And it is an elementary form of collective rather than purely individual intentionality, a precursor of and basis for saying “we” and becoming a subject of joint beliefs, plans, and values. Or so I shall argue.

Before outlining how I will go about this, let me say a few more preliminary words about my use of such notions as ‘representation’ and ‘intentional content’ because it is easily misunderstood in the present intellectual scenery. Many theorists nowadays recoil from these notions because they are associated with a cognitivist outlook that interprets representations as being essentially symbolic and having a formal, syntactic structure. They find this outlook inadequate and too intellectualist at least for basic minds (e.g. Hutto and Myin 2012) and therefore try to understand elementary forms of intentionality without appealing to content at all, for example in purely interactionist terms. To avoid misunderstandings, let me just say here that I reject the cognitivist myth of an arsenal of non-conscious rules and representations across the board – not just for basic minds – and that I think of mind as consisting entirely in states of consciousness or experience and corresponding dispositions to be in such states (compare Schmitz 2012). However, it would be awkward to try to do without the word ‘representation’ and its cognates, and it is positively a mistake to try to do without the notion of content. Intentionality without content does not make sense. Whenever there is intentionality in experience, there must be an answer to the question of how the world is – correctly or incorrectly – experienced as being and that answer specifies the content of that intentionality. When somebody feels tempted or even compelled to deny this, this shows that they must associate more with “content” than this and likely reveals that they have an intellectualist, language-centric notion of content. There is also another – mostly epistemological – set of motivations for rejecting the notion of content for the experiences of perception, action and joint attention, associated with so-called ‘relational’ accounts of experience that I will briefly address below.

In the following section I will now first more sharply delineate the phenomenon of joint attention based on how it is understood in the literature of developmental psychology. I will then sketch a framework for understanding the intentional structure of joint attention that introduces the concept of mode representation, which is based on the idea that in taking up postures, their subjects do not only represent what, for example, they perceive, believe, or intend, but also their own position – through what I call ‘attitude’ or ‘position’ mode – and themselves – through what I call ‘subject mode’ – vis-à-vis the relevant state of affairs. In the third section I will characterize the subject mode experience of joint attention based on findings in the literature of developmental psychology and introduce the PAIR-account of joint attention as a pragmatic and affective intentional relation. I will then argue that the mode account can explain an additional result from that literature, namely that infants understand their co-attenders better than people whose interactions with an object they merely observe. The fourth section addresses the question whether in joint attention episodes we understand the mental states of our co-attenders or perhaps even perceive them. I argue that in joint attention we understand intentional relations, but at a level that is prior to the differentiation of mind and body and criticize some other accounts as remaining too much in the grip of the dualism of mind and body.

2. Joint attention, content, and mode

The most fundamental fact about jointly attending to something is that the jointness is not a matter of *what we attend to*, but of attending *with somebody* (Campbell 2002, chap. 8). The main goal of this paper is to argue that the best way to understand this fact is through the notion of mode representa-

tions, through which the co-attenders experience themselves as co-subjects of a shared position with regard to the object of their attention. By contrast, and in accordance with the biases noted above, some have tried to understand joint attention and similar phenomena like common knowledge purely in terms of what individual subjects attend to or what they know. But it is hard to see how from such a perspective joint awareness can be distinguished from mere mutual awareness, from which it seems clearly different – though some highly technical attempts to accomplish this have been made in the literatures on common knowledge and joint attention (Lewis 1969; Peacocke 2005; Schiffer 1972).

Here is a scenario which I think shows why any attempt to treat joint attention as a merely perceptual, purely cognitive phenomenon must fail and cannot distinguish mere mutual from genuinely joint attention (with apologies for its homicidal character). 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, he will have the time to get his gun out and shoot the politician; otherwise the bodyguard could shoot her first. So these two are attending to the same object, they are mutually aware what the other is attending to and there is a causal relation between the direction(s) of their attention(s) – as has been suggested by some as a condition in an analysis of joint attention (as discussed by Campbell 2002, 162f). Still, it seems clear that this is not an instance of two people attending to something jointly. Again, mutuality is not the same as jointness. How do we get to jointness? We need to add a prosocial motivation and at least a disposition for joint action.

This suggestion is in accordance with proposals in developmental psychology, where a prosocial motivation to share an object, even to share it for the sake of sharing, is taken to be criterial for joint attention (Carpenter and Liebal 2011), and is often also thought to be unique to humans: even primates do not go around pointing out interesting things to each other, as humans do all the time. Joint attention episodes, which are usually taken to begin at around 12 months of age, often have a tripartite structure of (1) initiation by getting the other's attention, followed by (2) a referential point to the object to be shared, before culminating in (3) a "sharing look", the comment on the object, which closes the triangle through an affectively charged meeting of minds. The affect can be sheer pleasure and excitement about the object; concern, for example in "social referencing" when an infant checks back with someone, often the caretaker, whether a situation is safe; puzzlement, eye-rolling, and many more. This sharing I take to be a joint communicative action, at least when it is successful. Note that the claim is not that sharing itself is the same as joint attention, but just that one can only be in joint attention mode when one is also disposed to a joint communicative or other joint action. Otherwise one would not be attending with the other person, but just to her as in our example above. This also means that the joint attention mode as such is neither solely theoretical nor practical, but rather has both theoretical, mind-to-world direction of fit as well as practical, world-to-mind direction of fit aspects.

To highlight the irreducibly practical, pragmatic aspect of joint attention, some prefer the term 'joint engagement' (e.g. Dow 2012; Hobson and Hobson 2011). However, while in philosophy and psychology attention is often treated as a purely perceptual phenomenon – a reflection of the theory-biased

pre-occupation with perception – the common sense understanding of ‘attend’ clearly also has the pragmatic meaning, for example, when we say that the nurse attended to the patient. I will therefore stick to the established terminology. Finally, to conclude the discussion of definitional matters, there are also more intellectual forms of joint attention, such as we-mode deliberation and discussion, say in a board meeting or in a seminar, but, in accordance with most writers on the topic, I will restrict the notion to the more elementary, non-conceptual, sensory-motor-emotional forms of attending jointly.

I agree with the Relational View of joint attention (Campbell 2002; Seemann 2011) insofar as it emphasizes its triangular nature; the distinctive way in which the co-attender figures in that relation; and its experiential character – the latter against views such as Peacocke’s who see it as conceptual level propositional attitude. However, I reject its disjunctivist refusal to embrace the notion of the intentional content of joint attention experiences. The Relational View tends to take up an external, third person perspective on joint attention, asking: what makes true a statement of the form that x and y are jointly attending to z ? Relations indeed make statements of this form true, but these relations only exist because of the contents of individual minds, and I think we want to know more about what the contents of these individual minds are. How do I need to experience the other and the relation in which we participate in order to experience him as a co-attender rather than as a mere object of attention? This cannot be answered just in terms of saying that we are both constituents of an experiential joint attention relation. We need to characterize the individual experiences further, and to do that we need to appeal to the notion of the intentional content of the co-attenders’ minds as that in virtue of which they participate in these relations.

Moreover, that the existence of the relation depends not only on external facts, but also on the contents of individual minds further means that it can fail to obtain because the content of just one mind is not appropriate. That is, there can be and sometimes are illusory experiences of jointness, as when you turn to me excitedly to share something – only to discover that my attention has wandered away from the movie you were experiencing as an object of joint watching. So we need a notion of experience and an understanding of intentionality that allows us to locate experiences and intentional contents in the minds of individuals in the good cases where we do attend jointly, as well as in the bad cases of illusory experiences of jointness. And only intentionalism can provide such a notion and such an understanding. The Relational View only allows that we can have a false belief about joint attention experience, because it makes the existence of individual joint attention experiences dependent on the relation (Campbell 2002). But it seems wrong that the contents of our minds are so dependent on one another and that we can only err in thought about joint attention experience rather than in that experience itself. So we need to distinguish relational readings of ‘experience’, by means of which we report intentionally successful experiential relations – such as when we say we experienced the scenery – from ones where we merely say that we had an experience *as if* there was such a scenery present, or that our experience was illusory or hallucinatory.

Disjunctivism is chiefly motivated through epistemological concerns. The disjunctivist is worried that if we allow that there are intentionally contentful experiences in the bad as well as in the good cases, we would need to start our epistemological reflections from this place of subjective experience common between the good and bad cases. He is further worried that content

would become the object of our intentionality, intervening between mind and world and blocking our access to the latter, as it were – rather than enabling it, which I think is what it actually is doing. Suffice it to note here that neither follows from anything that has been said. This is not the place to discuss these epistemological worries, which apply to all forms of sensory experience and have nothing specifically to do with joint attention, except insofar as they are encouraged through the received notion of a propositional attitude, to which I now turn.

This received notion can be characterized through the following three claims:

1. The proposition is the object of the attitude.
2. The proposition is a truth-value bearer and yet at the same time the content of practical attitudes like intention as well as theoretical attitudes such as belief.
3. The intentional content of a propositional attitude is identical to that of the relevant proposition. The subject and the mode of the attitude make no contribution to content.

The idea that the proposition is the object of a propositional attitude, respectively the failure to clearly distinguish object and content, is also what encourages the notion that content is somehow between the subject and the world. (2) embodies the theory bias in the sense that it is claimed that something that – as a truth value bearer – must be theoretical, is contained in theoretical as well as practical attitudes. To the extent that it is thought that all intentional states are propositional attitudes – and historically this has often been thought – it also embodies the theory bias in the second sense I have distinguished. (3) is also related to the theory bias in that the traditional view is inspired by reports of propositional attitudes, where the subject and the type of attitude are seen from an external point of view as something that is the case – as opposed to how it seems from the subject's own point of view.

It would be hard to overstate the grip the traditional model has had and still has on the philosophical imagination. I have criticized it extensively elsewhere (Schmitz 2013) and do not have the space to repeat all these arguments, so I will be very brief. Propositions are not the objects of intentional states except in special circumstances such as, for example, when Californian voters make up their minds with regard to the propositions on their ballot. Rather the object of, for example, the belief that it is raining is just the corresponding state of affairs. Now suppose that the same state of affairs is also the object of an intention to make it rain, let us say on part of a general, whose troops have the corresponding capacity. As I have emphasized, on the traditional view even this practical attitude in some sense contains something from the theoretical domain. However, on reflection it is hard to make sense of this idea. It is not that the general predicts that it will rain on the basis of evidence in favor of this prediction. It is rather because his meteorologists tell him that it will not rain that he decides to make it rain! Nor is it plausible, some philosophers to the contrary, that intending is itself a form of believing. So I do not think that there is any sense in which the intending general takes a theoretical position vis-à-vis this state of affairs or that his practical attitude contains something theoretical. Rather the part of his attitude that represents the state of affairs (in this case, the action) that the belief is also about, is not yet a complete posture – that is, a bearer of a truth or other satisfaction value, a speech act or an intentional state. To get such a posture, we need to add the theoretical or practical position of the subject vis-à-vis the state of affairs.

(The mistake of (2) of the received view is to assume that the element common to different kinds of postures could be represented by something that itself has a satisfaction value.)

Now, the central claim I want to defend is that this position is itself represented. The subject represents and is aware not only of a state of affairs, but of his or her – or our – position vis-à-vis that state of affairs, or, as we can also say, her *relation* to that state of affairs. This awareness is typically backgrounded, the focus typically on the state of affairs, but it is still there. In order to be said to be intending, the general must have some awareness that he takes it upon him to bring about that it rains, that he takes practical responsibility for this, and so on. Correspondingly, in order to be said to believe something, a subject needs to be aware that the belief must be responsive to reality, that she takes theoretical, epistemic responsibility for the reality of the relevant state of affairs, and so on. Note, however, that the claim is not that the subject needs to apply a concept in taking up the posture, or need even to have such a concept. It is surely implausible that one should need the concept of belief to believe or the concept of intention to intend. Rather, it is sufficient that the subject has a sense of her position – in about the same way in which one can have a sense of somebody as a potential cooperation partner without having the concept of a cooperation partner – or that, as in joint attention experience, she experiences this position or relation.

The next step towards what I will call ‘subject mode’ is comparatively easy. For a subject cannot represent its relation to some state of affairs without representing itself. For example, I cannot represent my passive position vis-à-vis the objects of my perceptual states without representing myself. I experience these objects as impressing themselves *on me*. Put generally, the claim is that every posture also has an aspect of self-consciousness. We are never aware of objects (including state of affairs) from nowhere, as it were – and by nobody – but always situate them in relation to ourselves – spatially, temporally, causally, cognitively, conatively, and so on – and even in relation to our social and jointly taken positions. Self- and object-consciousness are inextricably linked, as Immanuel Kant argued already and many others such as Ludwig Wittgenstein, P. F. Strawson, Jean Piaget and Gareth Evans have since, often under Kant’s influence. And the most characteristic and fundamental use of ‘I’ is its use in subject position (Wittgenstein 1958), which may even be immune to error through self-identification (Evans 1982). That is, I can be wrong about whether it is my arm that I’m seeing, but not about the fact that it is me who is seeing the arm. In the terms I have introduced, the key to understanding self-awareness is to understand how it occurs in subject mode position, not as part of the what-content, of what I see, think, or am otherwise aware of. And just as there are pre-linguistic, non-conceptual and non-propositional forms of individual self-awareness (Bermúdez 1998), for example in perception and action, there are also corresponding forms of collective self-awareness in joint attention. I assume that non-conceptual forms of representation can be distinguished from conceptual level ones in terms of such properties as the absence of logical operations such as negation, the degree of differentiation of representational role – e.g. perceptual vs. linguistic structure – and of the abstractness / concreteness of representations, and the degree of context-dependence (Schmitz 2012).

To see how the proposed notion of mode representation has immediate benefits in understanding collective intentionality, consider common knowledge.

Standard analyses employ potentially infinite iterations of the following form:

x knows [that p];
y knows [that p];
x knows [that y knows that p];
y knows [that x knows that p];
x knows [that y knows that x knows that p];
y knows [that x knows that y knows that p];
...and so on...

This infinity is a result of the attempt to reduce the “we” and of treating mode as non-representational. Each ascription of knowledge to the other will produce a new knowledge position with regard to that knowledge, which is itself not represented – and thus is not in the square brackets – until the other then again takes up a new knowledge position, creating another unrepresented position, and so on. In contrast, on the current proposal, our subjects can just say or think e.g. “We know that it rains” to indicate their joint knowledge of the fact that it rains.

Note that on the present proposal there is no further position from which the theoretical or practical position of the individual or collective subject is represented. This also means that while the subject is represented as part of reality, it is not represented as something that is the case, from a theoretical point of view, nor is it represented from a practical point of view – at least when it is represented in subject mode, this is of course different when it is represented as part of the what-content. In subject mode representation the subject is aware of itself in relation to states of affairs, as occupying practical or theoretical positions with regard to them, not as an object of a further position.

Before we return to joint attention to try to elucidate it with the notion of mode representation, let me briefly summarize the argument of this section. I have argued against the Relational View of joint attention that we need a notion of the intentional content of joint attention experience to explain the specific way in which co-attenders experience each other and the relation they stand in with regard to what they attend to, as well as to account for cases of illusory joint attention. I have also criticized the traditional conception of intentional postures as propositional attitudes and proposed to replace it with a view according to which the structure of postures is such that: (1) one part of their content represents their object, which is (at least typically) a state of affairs, and which I call ‘state of affairs’ content or ‘what’ content because it is what is believed or intended; (2) another part is associated with the type of attitude of the subject, which I propose to understand in terms of the notion of the position of the subject vis-à-vis that state of affairs and which I therefore call “attitude mode” or “position mode” content; (3) yet another part of content represents the individual or collective subject of that position and is called ‘subject mode’ content.

3. Joint attention as pragmatic, affective and relational: the PAIR-account

But how do I experience the other so that, should she also experience me in this way, that makes us co-attenders rather than just mutual objects of awareness? I will now try to explain what it means to experience somebody as a co-subject rather than just an object of attention. In accordance with the thesis of an inextricable relation between self- and object-awareness, I will then argue

that the way I experience the other is also reflected in how I experience the world, or rather in how we jointly experience it. There are two main sources of inspiration for the idea that we experience others as co-subjects. One I have mentioned already, Wittgenstein's distinction between subjective and objective uses of 'I'. The other is the linguist Ronald Langacker's idea that we construe an entity subjectively when we construe it as part of or in relation to what he calls the ground, by which he means the speech situation with speaker and hearer, the immediate context, mental background, and so on (Langacker 1987). I will extend the notion of such subjective construction from linguistic, semantic content to the intentional content of experience, and accordingly I will speak of experiencing others as co-subjects or subjectively.

The basic idea here is that to experience something subjectively is to experience it as an extension of my (and thus as part of our!) perceptual or actional apparatus. Langacker uses the example of how you experience the glasses that you are wearing: normally your attention is not focused on them and you are mostly just aware of them (if at all) as something that improves your access to the world. Or think about how a tennis player experiences his racket as an extension of his actional apparatus, as improving his actional reach in the world. These examples can serve as metaphorical models for how in experiences of jointness we experience the other as a potential or actual partner for theoretical, epistemic as well as practical cooperation; as a source of information about the world and at the same time as somebody who will help and guide us; as somebody who draws my (our!) attention to new, exciting, interesting things and who I in turn want to show interesting things to; but also as somebody whom I can trust in a dangerous situation (e.g. social referencing). This is how to experience somebody as a co-subject of perception and action and thus a part of a shared, common ground rather than as a mere object of one's intentionality.

Again, this part of our experience is typically backgrounded; we are focused on the objects of our attention, not the co-subjects. When we focus on the other, we invariably construe her more objectively. We then look at her, not with her. (This is certainly at least partly what people who talk about 'objectification' have in mind.) The level of experience we are talking about here is also the level where we are attuned to others, resonate with them and are aligned with them in various ways, for example, with regard to mimic, gesture and posture. That we are more sympathetic to those who are attuned to us more or even imitate us with regard to such features and are more likely to respond positively to their wishes and requests is a well-known phenomenon often called the 'chameleon effect' (Chartrand and Bargh 1999).

Many insights into how others are experienced, understood and treated in joint attention come from studies that reveal the characteristic deficits autistic children show in this regard. I will present some of these results to show how they support the theoretical fruitfulness of the distinction between subject mode and object representation. Strikingly, when asked where a sticker should go, more than half of the children with autism, but not a single non-autistic child, never indicated the place by pointing to their own bodies rather than at the other's body (Hobson and Meyer 2005). This is a very vivid illustration of 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 her as somebody like oneself rather than as an object. Research by Peter Hobson and Jessica Hobson also shows a correlation between sharing looks and role reversals in joint action. They conclude that "the

results suggest that the mode of social perception that involves sharing looks [also] gives rise to self-other transpositions in imitation” (2011, 124). On the subject mode approach this can be explained as an instance of experiencing the other as a co-subject, as somebody who is like me, because people like me can perform the actions that I perform, and because I experience myself as forming a joint subject of action with the other, so that it does not matter so much who does what and we can switch easily between different roles in the pursuit of a common goal. Autistic children further engage much less in the kind of affirmative nodding people often display when listening to others, and only 3 of 16 children with autism showed a concerned look when the drawing of the tester was torn in a joint attention situation (Hobson et al. 2009). This shows that autism is also connected to deficits in experiencing the world with regard to other’s interests and concerns and thus supports the thesis of a deep connection between subject- and object awareness.

This interdependence of self- and object-awareness means that the jointness of joint attention is generally not only manifest in how the co-attenders experience each other, but also in how they see the world ‘with each other’s eyes’. So those who are bound together in a joint attention episode tend to experience the world as containing things that they want to draw the other’s attention to, but also that they might want to shelter him from; as good and interesting or bad and boring for the other, and as like or unlike things they have jointly experienced in the past. That is, joint attention means that the co-subjects are attuned and aligned with regard to cognitive and conative interests as well as with regard to their physical features and stances and that we often experience the world in relation to us and our common ground of shared interests and past experiences. A result from recent developmental psychology nicely illustrates and supports this point. Infants shared several toy ducks with one experimenter and then several teddy bears with another. They then entered a room with just one of the experimenters, in which a duck and a teddy bear picture were on the wall, and 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 intentional content rather than state of affairs content explains certain kinds of social understanding and social actions based on that understanding. For example, 14-months-old infants understood an ambiguous request by an adult on the basis of 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, Carpenter, and Tomasello 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” (2011, 397).

From the present perspective, what is most important about these experiments is that they show that the infants could understand the relation of familiarity between the adult and the old object and thus that the other object was new and interesting relative to it, as long as it was part of a shared familiarity, *a common ground established by joint attention*, but that they could not under-

stand it merely on the basis of observation. I think this strongly suggests that the affectively charged subject content rather than the object content explains the infants' understanding of the adult's request. They understood the adults relation to the familiar object as part of the attention relation they jointly experienced. This explains why they were able to cooperate with the adult by means of handing over the desired toy.

What's the common denominator of these findings? A slogan that I find useful here is that joint attention subject mode experience is a form of "like-me"-intentionality (Meltzoff 2007). I experience somebody as like me, when I feel that I can take on any role she can, facilitating role reversal; when I identify with her in an affectively charged way, am aligned with her, attuned to her and tend to affirm her postures; and when I refer to her through sameness, that is, through imitative forms of representation. As a mnemonic device, let us call this account of joint attention the PAIR account. 'P' because this intentional relation has an irreducibly practical, pragmatic aspect: joint attention essentially brings with it at least a disposition for joint communicative actions of sharing and normally also for other joint actions. The 'A' signifies that this relation is affectively charged and typically involves alignment with, attunement to, and affirmation of one's co-attenders. The 'I' indicates that this relation is intentional and involves identification with and imitation of one's co-attenders, including imitative styles of reference to them. Finally, the 'R' reminds us that this relation obtains in virtue of the representational contents in the subject's minds and fosters role reversal and reciprocity more broadly. In a nutshell, we can say that the PAIR account conceives of joint attention as a pragmatic, affectively charged intentional relation that obtains in virtue of mode content in the co-attenders' minds that binds them together as co-subjects. So far we have focused on subject mode content. In the next and final section of this paper I want to extend the account to position mode. Again, positions can also be thought of as relations, and at the sensory-motor-emotional level of joint attention we are dealing with actional, perceptual, and affective relations. So our question will be how the co-attenders understand these relations.

4. Understanding others and understanding minds

I'll restrict myself here to addressing one philosophically contentious question, namely, do the co-attenders understand each other as subjects of mental states? Do they perhaps even experience these mental states directly? These questions are pressing in the context of theorizing joint attention because, as we noted above, infants are usually assumed to start participating in joint attention around 12 months of age, while they only pass classical theory of mind tasks like the traditional false belief-test, which requires them to explain the action of a character who has false beliefs, at around 3 1/2 to 4 years. They are even more pressing if we subscribe to the traditional, theory-biased and individualist approach to understanding others described at the outset, because, naturally interpreted, it would suggest that children should first understand others' minds in an individualistic context geared towards explaining and predicting their behavior, and only then go on to engage in forms of collective intentionality based on shared propositional attitudes. And even if we accepted at face value the claim that newer versions of the false belief-test based on violation of expectation and anticipatory looking paradigms can push down the relevant age to 15 or even 13 months (Baillargeon, Scott, and He 2010) this would not quite solve the problem.

One possible solution is the suggestion that children engaged in joint attention can simply directly experience others' mental states. Some philosophers (e.g. Krueger and Overgaard 2012; Seemann 2008) have proposed this, even independently of our problem, arguing that we do possess a more direct route to understanding others than the already quite theoretical, reflective methods of theory-building and simulating others that the traditional approaches to social cognition of the theory-theory and simulation-theory appeal to. I think these philosophers are right that we do not need a theory and do not need to engage in simulation, at least not in any ordinary sense, to see what another person is seeing or doing, or that a person is happy or in pain. But does this mean that we perceive mental states? Can I see the other person's perceptual, actional, or affective experience? If we put the question in this way, this already seems much less plausible. Obviously, we need to be clear here what a mental state is, and this in turn of course is itself a contentious philosophical issue. However, it seems to me that the notion of internal, subjective states of experience or consciousness is still the core of our notion of a mental state. Leaving to the side some extremists who deny the reality of experience, this is true even for those, who, unlike me, are not ready to embrace the idea that the notion of such states (and of dispositions to be in such states), is in fact the only notion of mentality we need. And unconscious mental states are not really at issue in the cases of experiencing others that engage us here. So can we perceive others' states of consciousness? We certainly cannot experience them as objects in their heads and since – pace some forms of externalism – they are located in the head, that seems a good reason for denying that we can perceive them. But it could be objected that this is the wrong model for experiencing others' minds. "Look", one might say, "this fails because of the subjectivity of consciousness and because, as you have emphasized, our primary mode of access to others is as co-subjects, not as objects. So we experience their minds e.g. by following their gaze and jointly attending with them; by empathizing with their pain; and by cooperating with their actions." However, while this is how in the most basic cases we experience and understand other people, it does not show that we really experience their states of mind.

At this point surely the objection will be raised that this is just an artefact of the internalist and experientialist understanding of mind that I have adopted. Can't we just escape this conclusion by going for externalist, extended, enactive, embodied, or embedded conceptions of mind, or any combinations thereof? But this argument can be turned around. Adopting a different definition of mind as such can only lead to a merely verbally different result. We might be able to say then, for example, that we perceive mind in the sense that we perceive a larger complex of which mind in the sense of experience is just a part, but this still would not give sense to the notion that we perceive experience. (This is not to say that there is not anything valuable in the ideas that mind is essentially *related* to its environment, to the body, and to action and artefacts; there is, but it does not lie in redefining mind in these terms, but in helping us to both understand mind itself and how we understand mind through these relations.) But what is the alternative? Are we stuck with the traditional view that we just perceive the body and its behavior and relations to the world and can (at best) just infer mental states as internal states that explain and predict that behavior?

I think we are in a bind here. We are caught in the fly-glass of the dualism of mind and body. Here I am referring to the conceptual dualism (Searle 1992) in the sense of the idea that all our concepts or other representations can be

classified as being either representations of mind or of body, of mental or physical entities. In the literature, we can see many philosophers struggling with this dualism, saying things such as that we perceive mind “in the embodied behaviors” (Gallagher 2011, 298), or that we have a “minimal theory of mind” that “tracks” mental states without really referring to them (Butterfill and Apperly 2013). These philosophers sense the inadequacy of the dualism and are trying to move beyond it, but cannot quite shake it off. (In philosophy, we are often tempted to express our rejection of a dualism in its very terms.) I want to argue that we should simply reject it. There are many entities that cannot be happily classified as being either purely mental or purely physical, for example those making up institutional reality (Searle 1995). Is a university something mental or something physical? How about the property of being the capital of Croatia? Closer to our present concerns, is a person something mental or something physical? A person has properties of both kinds, but what is it as such?

I suggest we add the person to the list of entities that straddle the mental / physical divide, and also related entities such as animals. This is also strongly called for from a phenomenological point of view. It seems neither right that we experience other people and animals as mere bodies, nor as pure spirits, souls or minds. (It is tempting to say that we experience the others as embodied minds, or ensouled bodies, or something like that, but this seems to be another attempt to state a rejection of a dualism in its very terms.) In the same way, it seems also wrong to classify basic forms of affective, actional, and perceptual relatedness to the world as either purely mental or purely physical. In other words, I am suggesting that action, perception, as well as pain, happiness, and other affects should also be added to our list. This may sound implausible at first, but I think that on reflection it is actually quite intuitive that we begin with a gestalt-like understanding of, e.g. pain and perception, and only gradually, through a process of differentiation and abstraction, are able to distinguish pain experience from pain behavior and perceptual experience from the other components of the perceptual relation. Note that I’m not making a claim here about our use of words like ‘pain’ and that I am not trying to legislate that use. It is clear that pain experience is what most interests us about pain, and so it may well be the case that this is really what most people mean by ‘pain’ – though I would suspect the actual situation is much more complicated. The point is that at the level of basic forms of social intentionality such as joint attention we do not experience affect, action and perception as differentiated into their mental and physical aspects.

It might also be objected that I am now appropriating the relational understanding of perception and attention associated with views such as Campbell’s that I criticized earlier. But here we have to similarly distinguish between the perspective of the theorist and that of the subject of basic experience that we are trying to understand. There may be some similarities between how perception and attention are experienced at the basic level and the kind of deliberate primitivism of the Relational View, but the deliberate refusal of the latter to distinguish object and content at the theoretical level surely cannot be justified by a similar lack of differentiation at the level of basic experience.

The claim then is that at the level of basic sensory-motor-emotional experience subjects experience and understand affective, actional, and perceptual intentional relations between people and objects. They experience what others perceive and their affective responses and goals. They also register such rela-

tions – to use Butterfill and Apperly's (2012) apt term – and can, for example, on this basis understand people's requests and what they might be excited about, as we saw above in the findings from Moll and her colleagues. But I do not think there is sufficient evidence that children at this stage already have an understanding that such relations obtain in virtue of representational mental states that may or may not agree with reality. Such understanding would require conceptual abilities such as the mastery of logical connectives like negation and the capacity to *confront* perspectives with reality (Moll and Meltzoff 2011) – for example that of the subject of a false belief to the effect that an object is in a certain container with the state of the world, in which it is *not*. It is not surprising that it is cognitively very demanding to keep two contradictory representations in mind at the same time in this way and to ascribe one to another subject and use it to explain and predict their actions, as traditional versions of the false belief task require. In contrast, I do not think that newer versions of the traditional tasks really test for the same kind of capacities, just for important precursors of them. That a child, as evidenced by increased looking times, is surprised when a subject does not look for an object in the place where the child last experienced the subject interacting with it and rather expects, as evidenced by anticipatory looking, that it should go to the old place, shows that the child has registered the intentional relation between subject and object and is habituated to and expects certain patterns of interaction between them. But it does not show that the child has a conception of false beliefs and thoughts and beliefs about them. We are familiar with all kinds of patterns that we have no concepts of and do not think about. For example, we are used to people aligning with us and attuning to us in interactions in certain ways and – as shown by the chameleon effect – this will influence how we feel about them. But that is not something that we normally, much less necessarily, are aware of and register in thought. Our surprise at deviations from such patterns is not evidence that we have thoughts and beliefs about them, but often what gets us to think about them in the first place. So while such deviations are likely to be part of what gets a child to start thinking about the mind and its content, conclusive evidence that it has understood the representational mind is only provided by passing the traditional tasks. Before this developmental breakthrough the child experiences, registers and understands intentional relations between its co-subjects and objects, but not yet that these relations obtain in virtue of representational states with contents that may or may not match the world.

Conclusion

Let me conclude this paper by summarizing its main line of argument. Against the traditional, theory-biased and strongly individualistic approaches to understanding others I have in effect argued that the theoretical, explanatory and predictive stance of the radical interpreter can only emerge – through a process of differentiation – from more basic affective, actional, and perceptual modes of relatedness to others as we find in joint attention. I proposed to understand joint attention as a pragmatically and affectively charged intentional relation (PAIR), in which we are aligned with and attuned to others and experience them as co-subjects, who are like us and like to share things in the world with us. We also to some extent see the world with their eyes. I have argued that there is empirical evidence that it is actually this mode of joint attention rather than the stance of the observer that explains certain forms of

understanding others. And I have made a case that we best account for the intentional structure of joint attention by rejecting the traditional notion of propositional attitudes and replacing it with an understanding of postures according to which subject mode and attitude or position mode are also thought of as being representational. Finally, I proposed that in joint attention experience we understand others at a non-conceptual level prior to the differentiation of mind and body, which only occurs on the conceptual level.

The next step would be to show how, if we think of applying concepts of mental states as arising from such a process of differentiation rather than as inferring mind on the basis of observing mere behavior as on the received approach, we can get a new perspective on the traditional skeptical problem of other minds. But this must be left for another occasion.

References

- Baillargeon, Renée, Rose M. Scott, and Zijjing He. 2010. "False-Belief Understanding in Infants." *Trends in Cognitive Sciences* 14 (3): 110–118.
- Bermúdez, José Luis. 1998. *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.
- Butterfill, Stephen A., and Ian A. Apperly. 2013. "How to Construct a Minimal Theory of Mind." *Mind & Language* 28 (5): 606–637.
- Campbell, John. 2002. *Reference and Consciousness*. Oxford: Clarendon Press.
- Carpenter, Malinda, and Kristin Liebal. 2011. "Joint Attention, Communication, and Knowing Together in Infancy." In Axel Seemann (ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, Cambridge, MA: MIT Press, 159–182.
- Chartrand, Tanya L., and John A. Bargh. 1999. "The Chameleon Effect: The Perception–Behavior Link and Social Interaction." *Journal of Personality and Social Psychology* 76 (6): 893.
- Davidson, Donald. 1984. *Inquiries into Truth And Interpretation*. Oxford: Oxford University Press.
- Davidson, Donald. 2001. *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3*. Oxford: Clarendon Press.
- Dow, James M. 2012. "On the Joint Engagement of Persons: Self-Consciousness, the Symmetry Thesis and Person Perception." *Philosophical Psychology* 25 (1): 49–75.
- Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Clarendon Press.
- Gallagher, Shaun. 2011. "Interactive Coordination in Joint Attention." In Axel Seemann (ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, Cambridge, MA: MIT Press, 293–304.
- Hobson, Jessica A., Ruth Harris, Rosa García-Pérez, and R. Peter Hobson. 2009. "Anticipatory Concern: A Study in Autism." *Developmental Science* 12 (2): 249–263.
- Hobson, R. Peter, and Jessica Hobson. 2011. "Joint Attention or Joint Engagement? Insights from Autism." In Axel Seemann (ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, Cambridge, MA: MIT Press, 115–136.
- Hobson, R. Peter, and Jessica A. Meyer. 2005. "Foundations for Self and Other: A Study in Autism." *Developmental Science* 8 (6): 481–491.
- Hutto, Daniel D., and Erik Myin. 2012. *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, MA: MIT Press.

- Krueger, Joel William, and Søren Overgaard. 2012. "Seeing Subjectivity: Defending a Perceptual Account of Other Minds." *ProtoSociology (Online)* 47: 239–262.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Vol. I. Stanford, CA: Stanford University Press.
- Lewis, David. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.
- Liebal, Kristin, Tanya Behne, Malinda Carpenter, and Michael Tomasello. 2009. "Infants Use Shared Experience to Interpret Pointing Gestures." *Developmental Science* 12 (2): 264–271.
- Meltzoff, Andrew N. 2007. "'Like Me': a Foundation for Social Cognition." *Developmental Science* 10 (1): 126–134.
- Moll, Henrike, Malinda Carpenter, and Michael Tomasello. 2007. "Fourteen-Month-Olds Know What Others Experience Only in Joint Engagement." *Developmental Science* 10 (6) (November 1): 826–835.
- Moll, Henrike, and Andrew N. Meltzoff. 2011. "Joint Attention as the Fundamental Basis of Understanding Perspectives." In Axel Seemann (ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, Cambridge, MA: MIT Press, 392–412.
- Peacocke, Christopher. 2005. "Joint Attention: Its Nature, Reflexivity, and Relation to Common Knowledge." In Naomi M. Eilan, Christoph Hoerl, Teresa McCormack, and Johannes Roessler (eds.), *Joint Attention: Communication and Other Minds*, Oxford: Oxford University Press, 298–324.
- Schiffer, Stephen R. 1972. *Meaning*. Oxford: Clarendon Press.
- Schmitz, Michael. 2012. "The Background as Intentional, Conscious, and Nonconceptual." In Zdravko Radman (ed.), *Knowing without Thinking: Mind, Action, Cognition and the Phenomenon of the Background*, Basingstoke: Palgrave Macmillan, 57–82.
- Schmitz, Michael. 2013. "Limits of Intention and the Representational Mind." In Gottfried Seebass, Michael Schmitz, and Peter M. Gollwitzer (eds.), *Acting Intentionally: Individuals, Groups, Institutions*, Berlin: Boston: DeGruyter, 57–84.
- Searle, John R. 1992. *The Rediscovery of the Mind*. MIT Press.
- Searle, John R. 1995. *The Construction of Social Reality*. Simon and Schuster.
- Seemann, Axel. 2008. "Person Perception." *Philosophical Explorations* 11 (3): 245–262.
- Seemann, Axel. 2011. "Joint Attention: Toward a Relational Account." In Axel Seemann (ed.), *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, Cambridge, MA: MIT Press, 183–202.
- Wittgenstein, Ludwig. 1958. *The Blue and Brown Books*. Oxford: Blackwell.

Michael Schmitz

Zajednička pažnja i razumijevanje drugih

Abstract

U ovome radu kritiziram pretjerano individualističke i teorijski pristrane pristupe razumijevanju drugih te uvodim PAIR shvaćanje zajedničke pažnje kao pragmatične, afektivno nabijene intencionalne relacije. Tvrdim da ova relacija postoji kao intencionalni sadržaj u umovima sudionika u pažnji te da bismo trebali – nasuprot uvriježenom razumijevanju intencionalnih stanja kao propozicijskih stavova – priznati ono što nazivam ‘subjektni modus’ i ‘pozicijski modus’ intencionalnog sadržaja. Na temelju rezultata istraživanja razvojne psihologije, predlažem da ovaj subjektni modus sadržaja predstavlja sudionike u pažnji kao ko-subjekte, koji im nalikuju i koji su u najmanju ruku u mogućnosti djelovati zajedno s njima. Zaključujem tvrdeći da u zajedničkoj pažnji iskušavamo i shvaćamo afektivne, djelatne i percepcijske relacije na nekonceptualnoj razini koja prethodi razlikovanju uma i tijela.

Ključne riječi

zajednička pažnja, kolektivna intencionalnost, drugi umovi, intencionalizam, dualizam uma i tijela

Michael Schmitz

Gemeinsame Aufmerksamkeit und das Verstehen anderer

Zusammenfassung

In diesem Artikel kritisiere ich theorielastige und übertrieben individualistische Ansätze zum Verstehen Anderer und führe die PAIR-Auffassung gemeinsamer Aufmerksamkeit als einer pragmatischen, affektiv geladenen intentionalen Relation ein. Ich vertrete die Ansicht, dass diese Relation aufgrund der intentionalen Inhalte im Bewusstsein ihrer Subjekte existiert, und dass wir – im Gegensatz zur traditionellen Auffassung intentionaler Zustände als propositionale Einstellungen – anerkennen sollten, dass auch das, was ich als „Subjektmodus“ und „Positionsmodus“ bezeichne, intentionalen Gehalt hat. Basierend auf Forschungsergebnissen aus der Entwicklungspsychologie schlage ich vor, dass der Inhalt des Subjektmodus die Aufmerksamkeit-Teilnehmer als Mitsubjekte repräsentiert, die ihnen gleichen und zumindest geneigt sind, gemeinsam mit ihnen zu handeln. Ich schließe mit der Behauptung, dass wir in der gemeinsamen Aufmerksamkeit affektive, aktionale und perzeptuelle Beziehungen auf einer nicht-begrifflichen Ebene erfahren und verstehen, die der Differenzierung von Geist und Körper vorausgeht.

Schlüsselwörter

gemeinsame Aufmerksamkeit, kollektive Intentionalität, Verstehen Anderer, Intentionalismus, Dualismus von Körper und Geist

Michael Schmitz

L’attention conjointe et la compréhension des autres

Résumé

Dans cet article, je critique les approches de la compréhension des autres excessivement théoriques et individualistes, puis j’introduis l’explication PAIR de l’attention conjointe comme une relation intentionnelle pragmatique, chargée affectivement. J’affirme que cette relation a cours comme contenu intentionnel dans l’esprit des co-participants et que – à l’encontre de la compréhension habituelle des états intentionnels en tant qu’attitudes propositionnelles – nous devrions reconnaître ce que j’appelle contenu intentionnel en « mode sujet » et en « mode position ». Me fondant sur les résultats de la recherche en psychologie du développement, je propose que ce contenu en mode sujet représente les co-participants en tant que co-sujets, qui leur ressemblent et qui sont les moins disposés à agir conjointement avec eux. Je conclus en affirmant que dans l’attention conjointe nous vivons et comprenons les relations affectives, actives et perceptuelles à un niveau non-conceptuel préalable à la différenciation de l’esprit et du corps.

Mots-clés

attention conjointe, intentionnalité collective, autres esprits, intentionalisme, dualisme esprit-corps