

Attending, acting, and feeling together

Michael Schmitz

To cite this article: Michael Schmitz (11 Apr 2025): Attending, acting, and feeling together, *Philosophical Psychology*, DOI: [10.1080/09515089.2025.2489445](https://doi.org/10.1080/09515089.2025.2489445)

To link to this article: <https://doi.org/10.1080/09515089.2025.2489445>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 11 Apr 2025.



Submit your article to this journal [↗](#)



Article views: 119



View related articles [↗](#)



View Crossmark data [↗](#)

Attending, acting, and feeling together

Michael Schmitz

Institut für Philosophie, Universität Wien, Wien, Austria

ABSTRACT

In this paper, I argue that basic forms of collective intentionality such as those involved in attending, acting and feeling with others essentially involve experiencing and understanding others as co-subjects, that their content is nonconceptual, and that they represent co-subjects and their positions at a level that is prior to the mind-body differentiation.

ARTICLE HISTORY

Received 21 December 2023
Accepted 31 March 2025

KEYWORDS

Collective intentionality;
joint attention; joint action;
joint emotion; force/mode;
nonconceptual content

1. Introduction

What can philosophy contribute to the study of joint attention, action, and feeling, especially given that much work has already been done on these phenomena in psychology, particularly on joint attention (cf. Moll (2023) for a recent review)? A core task of philosophy is to critique conceptual frameworks and to propose and explore alternatives in light of the current state of research. A framework that is still pervasive in philosophy, psychology, and cognitive science is the framework of propositions and propositional attitudes. This framework exemplifies a way of thinking about meaning and mind that was originally developed to understand scientific theories. It arguably still exemplifies a bias for the theoretical both in the sense of a bias for intellectually demanding, high-level representations that are conceptually and propositionally structured (intellectualism) and for mind-to-world direction of fit postures (speech acts and attitudes) like assertions, beliefs, and states of theoretical knowledge over world-to-mind direction of fit postures like directives, intentions, and states of practical knowledge of what to do (cognitivism).

The tradition this framework belongs has also typically been reluctant to bring the subject of these postures into the picture. It's normally taken for granted that the subject and the type of posture as such make no contribution to content, which is typically taken to be propositional and conceptual and exhausted by *what* the subject e.g., believes or intends. If on the other

CONTACT Michael Schmitz  Michael.Schmitz@univie.ac.at  Institut für Philosophie, Universität Wien, Universitätsstr. 7, Wien 1010, Austria

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

hand the subject is thematized, it is assumed that it is something purely mental. This is also often assumed about about its perceptual, actional, and emotional states. Moreover, it is often thought that the way to understand the mind is by analyzing its contents into simple, atomic units. Finally, this picture also tends to be biased toward individual, solitary subjects. The social world is more typically seen as something that distracts from and biases people against accurate representations of reality rather than as something that enables such representations.

I think it is at least partly through the influence of this picture, that collective intentionality only became a topic in analytic philosophy starting in the 70s and 80s and that the basic forms of collective intentionality involved in jointly attending, acting, and feeling have only come to the fore quite recently (e.g., Campbell, 2002; Crone, 2021; Martens, 2021; Seemann, 2012; Wilby, 2023).

There are several reasons why the traditional picture makes it hard to properly understand collective intentionality. First, the fact that the representational content of a posture – a speech act or intentional state – on the received view of propositional attitudes is identified with its propositional content in the sense of what is asserted, believed, intended, and so on, makes it hard to understand how postures can have collective or plural subjects. On the received view, the subject and the type or mode of that posture do not make any contribution to its representational content. To be represented, it would have to be included in the propositional content of a *report* of that posture, a report that a certain subject e.g., asserted, believed or intended something. I think this partially explains why the still most popular account of common knowledge (e.g., Schiffer, 1972) can only represent it as a potentially infinite structure of recursive mindreading: I know that you know that I know that you know, and so on. Each new step will generate a new level of believing or knowing which is not itself represented, and representing it will create yet another level, and so on.

In contrast, the natural, common-sense way of representing common knowledge is just to say that *we* know something, for example, that it rains, where a subject is, so to speak, in self-conscious epistemic possession of a state of affairs (see Schmitz (2024) for more discussion). But to make sense of this at face value, we need at least two things: an account of postures where this *we* can be such a self-conscious *subject* of an attitude or speech act, and not just again part of the propositional content of what an individual believes; and an account of *we* that really makes sense of it as a collective, plural subject.

This second task is made more difficult by the tendency to think of the subject as something purely mental. If the individual I-subject is something purely mental, it seems like the collective we-subject must be, too, and that therefore it has to be something like like the dreaded group mind. It is,

I believe, fear of group minds that motivates both attempts to reduce the subject-we to *Is* such as Bratman's (2014) and Ludwig's (2016) and mode-accounts such as Tuomela's (e.g., Tuomela, 2013) and Searle's (1990) which accept the conceptual irreducibility of the we, but treat it as a mode in the sense of the received view at least insofar as it is not given a representational role as identifying a genuine we-subject. Searle further appeals to the idea of a pre-intentional, nonrepresentational background "sense of us".

While I think Tuomela and Searle are right that we-Intentionality, that thinking, attending, acting and feeling in a mode of identification with one's co-subjects, is irreducible to pure I-intentionality, it is hard to think of the we-mode as well as of a "sense of us" as being entirely nonrepresentational. Surely "we" and its cognates must represent oneself and one's co-subjects. Nor should we leave the idea of collective subjects to group mind mysticians – if they even exist. Collective subjects can be given a non-mysterious, rather straightforward interpretation: individual subjects are not something purely mental, but flesh-and-blood creatures or persons, and with "we", such subjects just mean themselves and other subjects *as related to them in certain ways*. The collective subject is not a group mind, nor is it free-floating relative to individuals, nor a fusion into which they disappear; it is rather a higher-level unity of such individuals and thus essentially a plural subject, a union of *Is*. It doesn't have a mind, but minds.

But how do these subjects represent each other as jointly attending, acting and feeling? When we try to answer this question, we run into further – and related – difficulties generated by the traditional picture. The picture invites us to always understand mental contents by analyzing them into simple, atomic constituents. The tendency to think that when we represent a subject as attending, acting or feeling with us, we must represent not only a purely mental subject, but also a purely mental state of attending, acting, or feeling, is an instance of this, as is the tendency to think that joint attention or perception can be understood independently of joint action and emotion. The tendency to think that experiencing and representing one another as jointly attending, acting or feeling must involve ascribing a purely mental state creates a well-known problem (e.g., Tollefsen, 2005) because at least according to one well-established view, which accepts passing classical versions of the false-belief test as criterial for understanding mind, children only acquire the capacity to ascribe mental states to others at around 3–4 years of age (Wellman et al., 2001), while they already engage in joint attention much earlier, from around 9–12 months, if not yet earlier (see Moll (2023) for discussion). How is this possible?

I have for some time argued that these difficulties with the traditional picture should lead us to break with it and embrace an alternative account (e.g., Schmitz, 2018). It is based on the following ideas. The first already applies at the individual level and is inspired by

a tradition of thinkers that goes back to Kant and includes Piaget, Strawson, Evans and contemporary philosophers such as Jose Luis Bermúdez, who think of self- and world- or object consciousness as essentially connected and like two sides of the same coin, or two poles – a subject and an object pole. I propose that in any posture, a subject is also always aware of its *position* vis-à-vis the state of affairs (SOA) or other object identified by the propositional content (as ordinarily conceived). Position can mean quite different things here. At the basic level, position is to be taken quite literally: it can mean the spatial position of the subject in the world that it experiences. The claim is that in perceptual experience, the subject always experiences itself as perceiving the world from a certain spatial position and the objects in the world as being spatially related to itself. It can also mean the subject's temporal position as indicated e.g., through tense. But we are especially interested here in position as it corresponds to the mode/force of a posture. For example, in perception the subject experiences the world as acting on itself, in action itself as acting on the world. In assertion I present myself as knowing what is the case, in direction as knowing what to do (Schmitz, 2022). In addition to content that represents the SOA or other object the posture is directed at (SOA or object content), there is therefore also content that represents the subject (subject content) and its position (position content).

The second idea is that collective intentionality can then be understood as an extension of the subject pole of individual intentionality. It occurs when we experience and represent others not as objects of our positions toward the world, but as their *co-subjects*: when, for example, we do not attend to them, but with them; when we think, reason, attend, act and feel in modes of identification with others, from the point of view or perspective of collectives.

The third idea is that this happens on different layers or levels of collective intentionality¹ and that these can be distinguished in terms of such parameters as the number of subjects involved (from joint attention dyads to nation states), the degree of role differentiation among them, and, most important for present purposes, the kind and structure of the relevant representational states and artifacts, including the kind and structure of their content. For purposes of rough orientation it is useful to distinguish three layers: the sensory-motor-emotional level of joint attention, action and feeling (JAAF), whose content is non-propositional and non-conceptual and that is the focus of this paper; the we-mode level, where we take propositional group positions in thought and speech, where we jointly believe, intend, assert, advise and adopt other propositionally and conceptually structured postures; and the institutional level, where individuals and groups take positions in their institutional roles such as being

a citizen or a committee, and which essentially requires writing and other forms of documentation.

The fourth idea is that one essential characteristic of the non-propositional and non-conceptual content at the JAAF level is that it is gestaltlike or dense and therefore less “analyzed” and differentiated than content that is propositional and conceptual. As I will argue in detail in this paper, this also means that at this level subjects experience themselves and their co-subjects as well as their own positions toward and relations to the world and those of their co-subjects’, at a level that is prior to the mind-body differentiation. A related point that I will also discuss is that the jointness of e.g., joint attention cannot be understood in purely perceptual terms alone, but requires an affective charge and at least a disposition for joint action, so that jointly attending, acting and feeling always come as a package.

In this paper, I will largely take for granted the first two ideas that postures generally have a subject and an object pole and that collective intentionality can be understood as an extension of the subject pole through co-subjective representation – except that I will apply these ideas to the JAAF level and if this application is plausible, it will also support this idea. In a companion piece to this paper (Schmitz, 2024) I have also already argued that jointly attending, acting and feeling are experiential relations to the world, but intentional relations that have to be understood in terms of content in the minds of the co-subjects. In a nutshell, experiential intentional relations require experiential states, and content by definition is just that feature of these states that puts their subjects into intentional contact with certain objects, but not others – and sometimes not at all. I will therefore also take for granted that the JAAF level should be understood in terms of content.

The present paper instead focusses on the following three tasks. In the next, second, section I will introduce what I call the “PAIR”-account of jointly attending, acting and feeling as pragmatic and affectively charged intentional relations, and in so doing I will show what co-subjective representation concretely means by applying this notion to a number of examples from developmental psychology. In the third section I will explain how I understand the notion of non-conceptual content and argue that the content that is partially constitutive of JAAF-relations is indeed non-conceptual in this sense. In the fourth and final section I will defend the idea that the non-conceptual representation of subjects and their relations to the world is prior to the mind-body differentiation.

2. Experiencing and representing others as co-subjects

What does it mean to represent somebody as a co-subject rather than as the object of a posture? I believe that this difference is a structural difference that is manifest in how co-subjects influence and guide one another. A co-

subject draws attention and invites imitation and identification. Its goals and feelings are contagious. Think about how our gaze and our attention are drawn when, for example, a group of people all look into a certain direction or run from something. This is of course not sufficient for jointness yet, but it illustrates how the experience of co-subjects has a structural, guiding role for our consciousness, for what we perceive, feel and do. We have a strong tendency to imitate our co-subjects, to align with them, to attune to their feelings, and to affirm them.

Subject content represents the subject as the subject, in various relations to the objects it is directed at. To experience another as a co-subject is to experience them as a potential or actual partner for such relations and for theoretical, epistemic or practical cooperation; as a source of information and as a guide to interesting things and as somebody who I in turn want to show interesting things to; as somebody whose feelings are appropriate to the situation at hand; as somebody who has my interests at heart and whom I can trust in a dangerous situation. For example, in developmental psychology the quality of the relation or attachment between an infant and its caretaker is measured in terms of its tendency to engage in what is known as “social referencing”, as when an infant exchanges a gaze with its caretaker when confronted with a dangerous or dangerous-seeming situation like a “visual cliff” (Gibson & Walk, 1960).

I believe that co-subjective relations are always affectively charged: certain feelings are an essential part of co-subjective experience. I will not try to give an exhaustive account of the feelings that can tie subjects together as co-subjects, and most likely such an account is not even possible, as any list of such feelings may be indefinitely extendible. However, two kinds of feelings that are certainly central are feelings of trust and feelings of joy. We need to trust those that we perceive, do and feel things together with, and we bond with those that we enjoy these things with.

To understand basic co-subjective relations and the JAAF-mode, I will in this section draw on earlier work (Schmitz, 2014) in which I introduced the PAIR account of joint attention as a perceptual-pragmatic and affectively charged intentional relation. It was put forward as an account specifically of joint attention, but one of its core point is that the jointness of joint attention cannot be understood independently of a pro-social motivation and at least a disposition for joint action.²

To make this point, I used the following scenario: “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.” (cf 2014, p. 238); two obvious mistakes were corrected).

The example is constructed to meet the criteria put forward by theorists who want to reduce joint attention to a relation of mutual perceptual awareness (including a causal condition formulated by John Campbell (2002), p. 162; see Schmitz, 2024 for more discussion). It is meant to make vivid the point that such mutual perception is not the same as joint perception or attention.³

As a mnemonic device, some key concepts of the PAIR account can be listed as follows:

Pragmatic, perceptual

Affectively charged, affirmation, attunement, alignment

Intentional, imitation, identification

Representational, reciprocity, role reversal

I will explain the PAIR account and these concepts here by illustrating their application to some findings drawn from the literature in developmental psychology. I think the fact that, as I will try to show, these examples can be naturally and straightforwardly interpreted on the present account, also supports this account. It should be kept in mind though that the account is a general philosophical framework rather than a theory that makes very specific predictions, and that its test is in whether it can make sense of the phenomena and inspire new ways of thinking and new research.⁴

Joint attention is “pragmatic” because it grounds at least a disposition for joint action – if it is not already in the service of joint action. This disposition is grounded in the affective charge of the joint experience, which must therefore be positively valenced. Jointness is therefore also essentially based on imitation and affirmation of our co-subjects and particularly basic embodied versions of these such as alignment and affect attunement – the cross-modal response to and affirmation of affect (e.g., Stern, 2018). That we are more sympathetic to those who imitate us more and are more aligned with and attuned to us, and more likely to act jointly with them, is a well-known phenomenon called the “chameleon effect” (Chartrand & Bargh, 1999).

The sensory-motor-emotional relations are intentional because their existence depends on the presence of intentional, representational content, including subject content, in the minds of the co-attenders. But, in keeping with the thesis of a close relationship between self- and object awareness, co-subjectivity is also manifest in how we experience the world and act in it when we are in JAAF-mode, even in moments when we are not actually experiencing our co-subject. For example, when running together with my wife, I’m much more sensitive to the presence of dogs, perceive them differently, have a different emotional response to them, and act differently to make sure we stay clear of them, because my wife is afraid of them.

Many insights into how others are experienced, understood and treated as co-subjects come from developmental psychology, particularly from studies that reveal the characteristic deficits autistic children show in this regard.⁵ For example, when asked where a sticker should go, no neurotypical child ever indicated the place where it should go by pointing to their own body rather than at the other's body, whereas more than half of the children with autism never did this (R. P. Hobson & Meyer, 2005). This vividly illustrates the difference between a co-subjective and an objectifying style of representation. To indicate the place by pointing to one's own body, through a form of imitation and identification, is to treat the other as somebody like me, a co-subject rather than a mere object.

Peter and Jessica Hobson also found a correlation between sharing looks and role reversals in joint action: "the results suggest that the mode of social perception that involves sharing looks [also] gives rise to self-other transpositions in imitation" (R. P. Hobson & Hobson, 2011, p. 124). The PAIR approach also explains this as an instance of experiencing the others as a co-subject, as somebody who is like me, because people who are like me can perform the same actions that I can perform. And when I form a joint subject of action with the other, it does not matter so much who does what and we can easily switch between different roles in the pursuit of a common goal. This is also confirmed by the literature on non-autistic children. As Tomasello notes, a study he did with Carpenter et al. (Tomasello et al., 2005) "found that after young children played one role in a collaboration, they could quickly switch to the other . . ." (2014, p. 40).

Autistic children also engage much less in the kind of affirmative nodding people often display when listening to others, and only 3 of 16 autistic children showed a concerned look when a tester's drawing was torn in a joint attention situation (J. Hobson et al., 2009) –, respectively, what would be a joint attention situation for neurotypical children, but appears to be less so for autistic children. This shows that autism is also connected to deficits in emotional identification with others in experiencing the world. It supports both that jointness essentially has an emotional component and that there is a deep connection between co-subjective awareness and object awareness.

This connection means that how a subject perceives the world, how it acts in it and feels about it, is partially shaped by the perceived theoretical and practical concerns of its co-subject(s) and their common ground and shared memories. For example, when people watch movies or experience other art together, they are much more likely to notice and point out things the other person would be likely to enjoy – or which it would not enjoy, in which latter case they might shield the other person from them, as in the dog example above. This is why

jointly experiencing something can be so rewarding, eye-opening and enriching.

An elementary example of how our past joint experiences shape our present ones and our communicative actions is again provided by an experiment in developmental psychology: infants shared several toy ducks with one experimenter and then several teddy bears with another. Then entering a room with just one of the experimenters that had a duck and a teddy bear picture on the wall, they were much more likely to point to the picture of the object they had earlier jointly experienced with the experimenter they were with (Liebal et al., 2009).

There is also some (defeasible) support for the idea that subject intentional content rather than SOA content explains certain kinds of social understanding and social and joint actions based on that understanding: 14-month-olds understood an ambiguous request by an adult on the basis of a joint attention episode that they shared, but not based on merely observing their otherwise identical interactions with the relevant objects. After the adult and the infant had jointly experienced two objects, and the infant had explored another, new, object by itself, it was able to correctly interpret an ambiguous request for “that one”, made with an excited expression by the adult, as referring to the new one. But 14-month-olds were not able to do the same in conditions where they 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” (2011, p. 397).

These experiments show that the infants could understand the relation of familiarity between the adult and the objects 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 understand it merely on the basis of observation. Under the assumption that what is individually observed is represented merely as a part of individual SOA content, this suggests that the affectively charged subject content plays an essential role in the infants’ understanding of the adult’s request. They understood the adult’s relation to the object as part of a co-subjective attention relation they jointly experienced, and this explains how they were able to cooperate with the request by handing over the desired toy.

Results by Liebal et al. (2009) show that joint action relations can also shape the interpretation of speech in similar ways: one-year-old infants interpreted an identical pointing gesture from an adult differently depending on whether they were in joint action mode with this adult or not.

3. JAAF mode intentionality as nonconceptual

What makes JAAF mode intentionality basic or minimal? Some have proposed that basic forms of collective intentionality such as joint attention differ from propositional attitudes like intention and belief in lacking content altogether. On John Campbell's Relational View (2002), joint attention just consists in a relation between the co-attenders and what they attend to. Enactivist views of joint attention instead understand it in terms of its relation to action. But both views reject the idea that the experience of joint attention has content. From the point of view of these theories, to ascribe content to it is to over-intellectualize this experience. However, it seems to me that this is because they assume an intellectualized notion of content as discussed above. I think it can be shown that all objections to applying the notion of content to individual or joint attention or perception are based on either assuming that content must be propositional and conceptual, or misconstruing content as a mental object, that – ironically – blocks direct access to the world rather than enabling it. I have elsewhere discussed objections along these lines both with regard to individual (Schmitz, 2019) and to joint perception (Schmitz, 2024), and have tried to characterize a minimal notion of content which I think is immune to them. On this minimal understanding, content is just that feature of postures that puts subjects into contact with certain objects, but not others. To embrace content in this sense, it seems to me, is just to be a realist about intentionality.

Here I will just focus on the positive task of characterizing the kind of content that is connected to basic forms of collective intentionality. I will try to make plausible that this content can be characterized in terms of the notion of nonconceptual content. This notion was initially developed to characterize the content of individual perception, but it has now also been applied to emotions (e.g., Tappolet, 2020), action (e.g., Pacherie, 2011) and self-consciousness (e.g., Bermúdez, 1998). I will try to show that it can also be applied to basic forms of collective intentionality.

Before I come to this, let me note that it is not so important here where exactly a line between nonconceptual and conceptual content should be drawn. This is bound to be a verbal question to some extent, and any such line will be fuzzy. What is important is to identify the relevant parameters. These parameters are not only useful for understanding the conceptual-nonconceptual distinction, but can also be used for characterizing levels or layers of intentionality more broadly. In the following I will introduce these parameters using examples from the domain of attending, acting and feeling together, thus trying to show that the notion of nonconceptual content can be fruitfully applied to that domain.

3.1. *Gestaltlike, holistic, dense, undifferentiated character*

Nonconceptual content is more gestaltlike, holistic and undifferentiated. For example, we visually experience objects as colored, but in vision color and shape are not differentiated in the way that they are at the linguistic, conceptual level. This point also applies to the social and collective domain. For example, a sharing look contains multitudes of feelings toward both object and co-subject – of joyful excitement about and relatedness to the latter – that one could spend ages conceptually differentiating. And as will be more extensively discussed later, the experience of the other subject and its perceptions, actions and feelings is at a level prior to the differentiation of body and mind.

3.2. *Context-dependence*

States with nonconceptual content are also more context-dependent than states whose content is conceptual. Context-dependence here means a causal kind of dependence – as opposed to the constitutive or definitional dependence the relationist has in mind. That is, the vividness, richness and immediacy of sensory experience typically cannot be sustained in the absence of the relevant objects. For example, sensory imagination will be deficient in these respects. Context-dependence can also be found in the domain of jointness. The rich, vivid and immediate experience of others is also dependent on their actual presence. We typically bond in face-to-face interactions. Though phones and video calls etc. complicate this picture, they can't replace face-to-face interaction across the board. If anybody had doubts about this, the experience of the Corona pandemic with its Zoom fatigue has certainly reinforced this observation.

The process of developing representations that are progressively context-independent is called “decontextualization” or “decoupling”. For example, thought and imagination are more context-independent than perception.

3.3. *Thought independence*

Evans (1982) introduced the idea of the “belief independence” of perception with references to examples such as the Müller-Lyer illusion: even when we know that the lines have the same length, they still seem to differ in length. This idea can be extended to the practical domain: an action can be triggered even when we intend not to perform that action or know not to perform it, as, for example, in so-called action slips and pathologies such as anarchic hand syndrome, where the person's hand may, for example, take food from another person's plate. In these kinds of cases we may speak of “intention independence”, and more generally, to cover both the theoretical and

practical domain, of “thought independence” – the relative independence of the conceptual level of theoretical and practical thought from the nonconceptual sensory-motor-emotional level. This independence can also be found in the collective domain. For example, a couple may fall back into patterns of destructive interactions, triggered by certain looks, gestures, or tones of voice, despite their best intentions. As members of a soccer team, we may know not to open up our defense too soon after conceding a goal, only to fall back into this behavioral pattern again and again. And we may also jointly undergo an illusion in spite of jointly knowing it to be one, sharing both illusion and knowledge with a sharing look or other communicative act.

3.4. Absence of logical operations

Logical operations require conceptual content. They operate on thoughts, not perceptual, actional or emotional experience. For example, there is no negation or disjunction in visual experience (Block, 2023). Nor is there, for example, in the joint emotional experience of enjoying the sunset together.

3.5. Representational role differentiation

Thought and speech that is propositional and conceptual is articulated into constituents which have distinct representational roles in the wholes of which they are parts. “Singular term” and “predicate” are logical terms for such roles, “subject”, “object”, “verb” and “adjective” grammatical ones. Higher levels of intentionality are inter alia characterized through a higher degree of representational role differentiation. So propositional and conceptual thought and speech show more role differentiation than sensory-motor-emotional experience, and written language shows even more than spoken language, as its grammatical structure tends to be more complex. One piece of evidence that conceptual thought is more differentiated than sensory-motor-emotional experience is that it comes in units such as sentences or propositions, because such units presuppose role differentiation. A unit is completed when all of its elements have been produced. In contrast, sensory-motor-emotional experience has the form of a continuous flow. This is certainly not to say though that it is not structured at all. Most importantly for present purposes, it has an attentional structure: there is a focus and a periphery of experience.

4. Experiencing and understanding our co-subjects’ positions⁶

How do we experience and understand the positions of others when engaging with them in episodes of attending, acting and feeling together? As we

noted above, this question turned out to be hard to answer for proponents of the traditional approach, which tries to understand collective intentionality solely in terms of propositional attitudes. This is because infants are usually assumed to start participating in basic forms of joint intentionality such as joint attention at around 12 months of age, if not even significantly earlier. But they only pass classical theory of mind tasks which test for their understanding of such “propositional attitudes” like the traditional false belief-test (Wellman et al., 2001), which requires them to explain the actions of a character with false beliefs, at around 3 1/2 to 4 years. 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 months (Onishi & Baillargeon, 2005) or even further (Surian et al., 2007), this would not quite solve the problem – especially in view of the just noted fact that the estimate for the emergence of joint intentionality is also conservative and may have to be revised downward, and perhaps significantly.

In any case, it is plausible that there are forms of experiencing and understanding other creatures and their perceptual, actional and affective relations to the world which are more basic than an understanding of belief which requires a child to *explain* what somebody does. One suggestion that has recently been pursued by philosophers (e.g., Krueger & Overgaard, 2012; Seemann, 2008), even independently of our problem, is that we can simply perceive what other creatures perceive, do and feel. It seems to me that these philosophers are right that we do not need a theory (in the sense of the “theory-theory” of social cognition) and that we do not need to engage in simulation (in the sense of the simulation theory of social cognition) to become aware of another creature as e.g., seeing a dog, closing a door, or being happy or in pain. But does this really mean that we can see or hear mental states? Can I see or hear another creature’s perceptual, actional or affective experience? Put in this way, this already seems much less plausible.

A crucial part of the issue here obviously is the question what a mental state is, itself a contentious philosophical problem. It seems to me that internal, subjective states of consciousness still form the core of our notion of a mental state. Leaving to the side those who deny the reality of consciousness, this is even true for those, who, unlike myself, are not ready to embrace the idea (Searle, 1992; Strawson, 1994) that the notion of such states – and of dispositions to be in such states – is actually the only notion of mentality we need. That is, I think that even those who believe in non-dispositional, occurrent unconscious mental states, must accept that states of consciousness or experience – I use these terms interchangeably here – are the paradigm mental states. If an argument for this is wanted, it can be that to even ask the question whether there are occurrent unconscious

mental states, we need to ask something like “Are there such states that are sufficiently like states of consciousness to also count as mental states?”. To simply start with a notion of an unconscious mental state I think is to beg that question. Moreover, unconscious mental states are not really at issue in the kinds of cases of perceiving others we are centrally interested in here.

Can we perceive others’ states of consciousness then? If we conceive of them as internal states located in organisms (more specifically in subjects’ heads), as I think we should, it seems plain that we cannot perceive them. And that’s not because they are private mental objects only accessible to their subjects. It is because they are subjective states that put subjects into perceptual contact with objects and thus not even the kind of things that can themselves be perceived. If we try to imagine perceiving them, we only end up imagining perceiving the brain or some other object – and this may tempt us to deny that experiences are brain states and located in the head. We can only get at the other’s state of consciousness in a co-subjective way, by imagining to experience the world from their point of view.

But what if we reject this internalist and experientialist understanding of mind and instead go for externalist, extended, enactive, embodied or embedded conceptions of mind, or any combinations thereof? However, again leaving to the side the denial that there is anything answering to this conception, that consciousness exists, adopting a different definition of mind can just lead to a verbally different result. We may then be able to say, 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 (something like that seems to be the view in e.g., Krueger and Overgaard), but this would still not give sense to the notion that we perceive experience in the sense used here.

It seems to me then that there is no point in redefining mind just to be able to say that we perceive it. Let me emphasize though that this does not mean that there is nothing valuable in notions such that mind is essentially *related* to the environment, to the body, to action and even to artifacts. I believe that there is, but that it does not lie in redefining mind so as to literally incorporate things external to it, but in understanding both mind itself and our understanding of mind in terms of its essential relatedness to the world, to action and to the body. Conversely, what gets us into certain traditional problems about mind I believe is not per se the thought that it consists of internal, experiential states, but that these states can be grasped and understood independently of 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 the presence of mental states as internal states that explain and predict that behavior? I think we are caught in the fly-glass of the dualism of mind and body here. I have in mind the conceptual dualism

(Searle, 1992) in the sense of the notion that all of our concepts (and other representations) represent their objects as being either mental or physical. In the literature, we can find many philosophers struggling with this dualism. They say such things as that we perceive mind “in the embodied behaviors” (Gallagher, 2011, p. 98), or that we have a “minimal theory of mind” that “tracks” mental states without really referring to them (Butterfill & 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 the dualism. There are many entities that cannot be happily classified as either purely mental or purely physical, for example those making up institutional reality (Searle, 1995). Is a university something mental or something physical? It has both physical properties like buildings, but also intentional, mental properties like purposes, but there is no clear reason why as such it should be either physical or mental. And this also seems to be true of entities like persons or creatures. I suggest that we therefore think of all these entities as straddling the mental/physical divide. This is also called for from a phenomenological point of view. It neither seems right to say that we experience other people or other creatures as mere bodies, nor as pure spirits, souls or minds. It may again be tempting to say that we experience others as embodied minds or ensouled bodies, or something of this kind, but this seems to be yet another attempt to state a rejection of the dualism in its very terms. It is therefore better to just say that we perceive them as people or creatures and that this experience is prior to the conceptual mind-body dualism.

I further want to suggest that we take the same attitude toward perceptual, actional and affective relations toward the world. That is, the most basic ways of experiencing and understanding these relations are also prior to the mind-body differentiation. For example, on the basic level subjects experience somebody as perceiving a tree without yet being able to differentiate or analyze the subject’s contribution to this relation into its bodily state and behavior on the one hand and an internal experiential state with a content distinct from its object on the other. Or we may experience and understand what a creature is doing as reaching for food or closing the door without yet being able to analyze this action into its component of a contentful mental state such as a motor (or other kind of) intention and the creature’s behavior. The understanding may, for example, just be manifest in the ability to complete the action or to help the other perform it in an appropriate way. And we may experience somebody as fearful without yet being able to distinguish the experience of fear and fear behavior. The understanding of fear at this basic level may just be manifest in becoming fearful oneself, in identifying the object of the fear or at least looking for

a possible cause of the fear, or in trying to comfort the fellow creature and to relieve or assuage their fear.

On this view then, there is a basic level of understanding others' perceptual, actional and affective relations to objects in the world that is prior to the differentiation of subjects' mental and physical properties, of e.g., distinguishing perceptual, actional and affective experience from perceptual, actional and affective behavior, or the contents of these experiences from their objects. Such distinctions are only acquired through a process of differentiation triggered by cases such as those where the content of an illusion (or other false beliefs) lacks a matching object; where an intention fails to be executed; or where a subject insincerely expresses theoretical or practical attitudes, or feelings, it does not actually possess.

At the sensory-motor-emotional level subjects experience and understand perceptual, actional and affective relations between other creatures and the world. They also register such relations – to use Butterfill and Apperly's (2013) apt term – and can on that basis e.g., understand people's requests and what they might be excited about, as we saw above in the results from Moll and her colleagues. However, this does not mean that at this level subjects already have an understanding that such relations obtain in virtue of representational mental states that may or may not match reality. Such an understanding would require the ability to *confront* perspectives with reality (Moll & Meltzoff, 2011), that is, for example, to simultaneously hold in mind the thought that a creature believes that an object is in a certain container and that actually it is *not*. But this ability is an ability for conceptual thought, as can be shown by applying several of the criteria introduced earlier. First, it involves *decontextualization* or *decoupling* from the perceived and/or believed reality of the container's actual location. Such decoupling requires thought or at least imagination. Second, the confrontation of perspectives requires the thought that the container is *not* actually in a certain location, though it seems to the creature under consideration that it is. But logical operations such as negation are a hallmark of conceptual content. Third, negation in turn can only operate on something that has a propositional structure and thus displays a high degree of representational role differentiation.

In contrast, newer versions of the traditional false belief task that operate with looking time paradigms (Onishi & Baillargeon, 2005; Senju et al., 2009; e.g.; Southgate et al., 2007; Surian et al., 2007) do not really test for the same kinds of abilities, just for precursors of them.⁷ The capacities they test for are preconceptual rather than conceptual capacities according to the criteria given above. For example, that a child is surprised – as evidenced by increased looking times – when a subject does not look for an object in the place where the child last experienced it interact with it, and that it rather expected – as evidenced by anticipatory looking – that it should have gone to the old place, shows that the child has registered the intentional

relation between subject and object, and that it is habituated to and thus familiar with certain patterns of interaction between subjects and objects. That is why the child expects the subject to conform to these patterns and is surprised when it doesn't. But this does not show that the child has a concept of these patterns, much less that it has a concept of false belief and the ability to form thoughts and beliefs about it. We are familiar with all kinds of patterns that we have no conception of and cannot think about.⁸ For example, we are used to people aligning with us gesturally and mimetically and imitating us in the way that the concept of the "chameleon effect" captures and would be surprised if they didn't, would experience them as being unresponsive and cold. And such an experience may get us to think about this phenomenon and perhaps form a conception of it; but our surprise does not show that we already possess such a conception.

In the same way, the surprise shown by a child about deviations from certain patterns of interaction between subjects and objects may be part of what triggers thoughts that eventually lead to a conceptual separation between the contents of mind and their objects, but is not yet sufficient evidence that it has formed the corresponding concepts. It seems to me such evidence is still only provided by the more traditional theory of mind tasks. Before it masters these tasks, the child experiences, understands and registers intentional relations between its co-subjects and objects, but it does not yet understand that these relations obtain in virtue of intentional states with contents distinct from their objects. Nor is it necessarily able to separate the behavioral expressions of these mental states from the mental states themselves. It seems to me that therefore the understanding that it has at this level is best thought of as being prior to the mind-body differentiation and that this differentiation requires conceptual thought.

5. Conclusion

I have argued that basic forms of collective intentionality such as those involved in attending, acting and feeling with others essentially involve experiencing and understanding others as co-subjects, that their content is nonconceptual and that they represent co-subjects and their positions at a level that is prior to the mind-body differentiation.

Notes

1. For a general discussion of layers or levels see List (2017). The concept of layered representation is familiar in cognitive science, see e.g., Green (2017). A layered account of infant intersubjectivity is given in Stern (2018) and of collective intentionality in Tomasello (2014), who distinguishes joint and collective intentionality, which, however,

- for him, both are forms of thought. A comparison of Tomasello's approach to the present one is beyond the scope of this paper, but can be found in (Schmitz, 2016).
2. The essential connection between joint attention and joint action is also emphasized by Tomasello: "Joint actions, joint goals, and joint attention are thus of a piece, and so they must have co-evolved together" (2014, p. 39).
 3. A similar, but less homicidal, example is given by Harder (2022). Again compare Schmitz (2024) for more extensive discussion.
 4. For an approach in psychology that is close in spirit to the present one, see (Shteynberg, 2015; Shteynberg et al., 2023).
 5. It should be emphasized though that, as the data to be reported also reflect, there is of course a lot of individual variation between autistic people, and that some can learn to more naturally adapt to the rhythms and patterns of the dance of sensory-motor-emotional interactions with more neurotypical people, though it takes effort. Thanks to an anonymous referee for the journal for these points.
 6. This section draws on material from (Schmitz, 2014).
 7. Also see the critical discussion in (Buttelmann et al., 2009).
 8. Part of the issue of course is here what a concept is, which is partly a verbal issue. I assume that a concept is essentially a constituent of thought and speech. Others use weaker notions. I think the criteria introduced above can be useful for describing and resolving such differences, but showing this is beyond the scope of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Bermúdez, J. L. (1998). *The paradox of self-consciousness*. MIT Press.
- Block, N. (2023). *The border between seeing and thinking*. OUP Usa.
- Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford University Press.
- Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112(2), 337–342. <https://doi.org/10.1016/j.cognition.2009.05.006>
- Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28(5), 606–637. <https://doi.org/10.1111/mila.12036>
- Campbell, J. (2002). *Reference and consciousness*. Clarendon Press.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893–910. <https://doi.org/10.1037/0022-3514.76.6.893>
- Crone, K. (2021). Foundations of a we-perspective. *Synthese*, 198(12), 11815–11832. <https://doi.org/10.1007/s11229-020-02834-6>
- Evans, G. (1982). *The varieties of reference*. Clarendon Press.
- Gallagher, S. (2011). Interactive coordination in joint attention. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 293–304). MIT Press.

- Gibson, E. J., & Walk, R. D. (1960). The “visual cliff”. *Scientific American*, 202(4), 64–71. <https://doi.org/10.1038/scientificamerican0460-64>
- Green, E. J. (2017). A layered view of shape perception. *British Journal for the Philosophy of Science*, 68(2), 355–387. <https://doi.org/10.1093/bjps/axv042>
- Harder, R. (2022). Joint attention and communication. *Inquiry, an Interdisciplinary Journal of Philosophy*, 67, 10):3796–3834.
- Hobson, J., Harris, R., García-Pérez, R., & Hobson, R. P. (2009). Anticipatory concern: A study in autism. *Developmental Science*, 12(2), 249–263. <https://doi.org/10.1111/j.1467-7687.2008.00762.x>
- Hobson, R. P., & Hobson, J. (2011). Joint attention or joint engagement? Insights from autism. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 115–136). MIT Press.
- Hobson, R. P., & Meyer, J. A. (2005). Foundations for self and other: A study in autism. *Developmental Science*, 8(6), 481–491. <https://doi.org/10.1111/j.1467-7687.2005.00439.x>
- Krueger, J. W., & Overgaard, S. (2012). Seeing subjectivity: Defending a perceptual account of other minds. *ProtoSociology (Online)*, 47, 239–262.
- Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, 12(2), 264–271. <https://doi.org/10.1111/j.1467-7687.2008.00758.x>
- List, C. (2017). Levels: Descriptive, explanatory, and ontological. *Nous*, 53(4), 852–883. <https://doi.org/10.1111/nous.12241>
- Ludwig, K. (2016). *From individual to plural agency: Collective action I*. Oxford University Press.
- Martens, J. H. (2021). Habit and skill in the domain of joint action. *Topoi*, 40(3), 663–675. <https://doi.org/10.1007/s11245-020-09732-z>
- Moll, H. (2023). What we do and don’t know about joint attention. *Topoi*, 43(2), 247–258. <https://doi.org/10.1007/s11245-023-09961-y>
- Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science*, 10(6), 826–835. <https://doi.org/10.1111/j.1467-7687.2007.00615.x>
- Moll, H., & Meltzoff, A. N. (2011). Joint attention as the fundamental basis of understanding perspectives. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 392–412). MIT Press.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255–258. <https://doi.org/10.1126/science.1107621>
- Pacherie, E. (2011). Nonconceptual representations for action and the limits of intentional control. *Social Psychology*, 42(1), 67. <https://doi.org/10.1027/1864-9335/a000044>
- Schiffer, S. R. (1972). *Meaning*. Clarendon Press.
- Schmitz, M. (2014). Joint attention and understanding others. *Synthese Philosophica*, 29(2), 235–251.
- Schmitz, M. (2016). A history of emerging modes? *Journal of Social Ontology*, 2(1), 87–103. <https://doi.org/10.1515/jso-2015-0054>
- Schmitz, M. (2018). Co-subjective consciousness constitutes collectives. *Journal of Social Philosophy*, 49(1), 137–160. <https://doi.org/10.1111/josp.12228>
- Schmitz, M. (2019). The good, the bad and the naive. In C. Limbeck-Lilienau & F. Stadler (Eds.), *The philosophy of perception* (pp. 57–74). De Gruyter.
- Schmitz, M. (2022). Force, content and the varieties of unity. In G. Mras & M. Schmitz (Eds.), *Force, content and the unity of the proposition* (pp. 71–90). Routledge.
- Schmitz, M. (2024). Joint attention: The PAIR account. *Topoi*, 43(2), 271–282. <https://doi.org/10.1007/s11245-024-10035-w>

- Searle, J. (1995). *The construction of social reality*. Free Press.
- Searle, J. R. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 401–416). MIT Press. http://people.su.se/~snce/texter/Coll_int.pdf
- Searle, J. R. (1992). *The rediscovery of the mind*. MIT Press.
- Seemann, A. (2008). Person perception. *Philosophical Explorations*, 11(3), 245–262. <https://doi.org/10.1080/13869790802239201>
- Seemann, A. (2012). Joint attention: Toward a relational account. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (p. 0). The MIT Press. <https://doi.org/10.7551/mitpress/8841.003.0010>
- Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in asperger syndrome. *Science*, 325(5942), 883–885. <https://doi.org/10.1126/science.1176170>
- Shteynberg, G. (2015). Shared attention. *Perspectives on Psychological Science*, 10(5), 579–590. <https://doi.org/10.1177/1745691615589104>
- Shteynberg, G., Hirsh, J. B., Wolf, W., Bargh, J. A., Boothby, E. J., Colman, A. M., Echterhoff, G., & Rossignac-Milon, M. (2023). Theory of collective mind. *Trends in Cognitive Sciences*, 27(11), 1019–1031. <https://doi.org/10.1016/j.tics.2023.06.009>
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18(7), 587–592. <https://doi.org/10.1111/j.1467-9280.2007.01944.x>
- Stern, D. N. (2018). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. Routledge.
- Strawson, G. (1994). *Mental reality* (Vol. 46, issue 184, p. 414). MIT Press.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, 18(7), 580–586. <https://doi.org/10.1111/j.1467-9280.2007.01943.x>
- Tappolet, C. (2020). Emotions inside out: The content of emotions. In C. Demmerling & D. Schröder (Eds.), *Concepts in thought, action, and emotion: New essays*. Routledge.
- Tollefsen, D. (2005). Let's Pretend!: Children and joint action. *Philosophy of the Social Sciences*, 35(1), 75–97. <https://doi.org/10.1177/0048393104271925>
- Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.
- Tomasello, M., Carpenter, M., & Hobson, R. P. (2005). The emergence of social cognition in three young chimpanzees. *Monographs of the Society for Research in Child Development* 70(1), i–152.
- Tuomela, R. (2013). *Social ontology: Collective intentionality and group agents*. Oxford University Press.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655–684. <https://doi.org/10.1111/1467-8624.00304>
- Wilby, M. (2023). The form and function of joint attention within joint action. *Philosophical Psychology*, 36(1), 134–161. <https://doi.org/10.1080/09515089.2022.2039384>