



# Joint Attention as the Base of Common Knowledge and Collective Intentionality

Axel Seemann<sup>1</sup>

Accepted: 16 January 2024 / Published online: 22 February 2024  
© The Author(s) 2024

## Abstract

I argue that joint attention solves the “base problem” as it arises for Schiffer’s and Lewis’s theories of common knowledge. The problem is that an account is needed of the perceptual base of some forms of common knowledge that gets by without itself invoking common knowledge. The paper solves the problem by developing a theory of joint attention as consisting in the exercise of joint know-how involving particular and sometimes distal targets and arguing that certain joint perceivers can always have a minimal form of propositional common knowledge about the location of these targets. On such a view, perceptual common knowledge is based on the experience of a process that is maintained by way of perceivers’ exercise of an object-involving form of joint know-how. Some reductive theories of collective intentionality require that agents’ intentions and subplans are common knowledge (or “out in the open”) between them. For these theories the base problem arises again. The enacted theory of joint attention can solve the problem. The argument is exactly parallel to the common knowledge case. The openness of joint agents’ intentions and meshing subplans is explained by appeal to their practical knowledge of how to maintain the process by way of which they pursue the collective intention. They can then make this knowledge explicit by linguistic communication. When they succeed in communicating knowledge of their meshing subplans as pursued in a joint action context, they necessarily have this knowledge in common. For theories of collective intentionality that include a common knowledge condition, the experience of participating in a perceptually constituted joint action provides the base that renders harmless the regress that otherwise threatens reductive analyses.

**Keywords** Joint attention · Common knowledge · Perceptual openness · Joint know-how · Interaction · Social space

## 1 Introduction

Joint attention has been credited with facilitating a vast range of social and cognitive functions in human ontogeny, ranging from the understanding of other minds<sup>1</sup> and social and spatial perspective-taking (Moll & Kadipasaoglu 2013; Moll & Meltzoff 2011) to language acquisition (Tomasello 2014) and the comprehension of perceptual objects as publicly accessible (Seemann 2022). It also plays an obvious role in adult human cooperation and interaction. In this paper I argue that the reach of joint attention is even more

comprehensive than these suggestions allow: it solves the “base problem” as it arises for some classic theories of perceptual common knowledge. The problem is that an account is needed of the perceptual base of some forms of common knowledge that gets by without itself invoking common knowledge. The paper solves the problem by developing a theory of joint attention as consisting in the exercise of joint know-how involving particular and sometimes distal targets and arguing that certain joint perceivers can always have a minimal form of propositional common knowledge about the location of these targets. On such a view, perceptual common knowledge is based on the experience of a process that is maintained by way of perceivers’ exercise of an object-involving form of joint know-how. Some reductive theories of collective intentionality (e.g., Bratman 1992, 1999) require that agents’ intentions and subplans are common knowledge (or “out in the open”) between them. For these theories the base problem arises again. The enacted

---

Thanks go to both of my reviewers for their extremely detailed and constructive comments.

---

✉ Axel Seemann  
aseemann@bentley.edu

<sup>1</sup> Department of Philosophy, Bentley University, Waltham, USA

<sup>1</sup> See e.g. the essays collected in Eilan et al. (2005).

theory of joint attention can solve the problem. The argument is exactly parallel to the common knowledge case. The openness of joint agents' intentions and meshing subplans is explained by appeal to their practical knowledge of how to maintain the process by way of which they pursue the collective intention. They can then make this knowledge explicit by linguistic communication. When they succeed in communicating knowledge of their meshing subplans as pursued in a joint action context, they necessarily have this knowledge in common. For theories of collective intentionality that include a common-knowledge condition, the experience of participating in a perceptually constituted joint action provides the base that renders harmless the regress that otherwise threatens reductive analyses.

The paper has three parts. I begin with a discussion of Schiffer's (1972) and Lewis's (1969) classic analyses of common knowledge and argue that they give rise to what I call the "base problem". The problem is that insofar as their analyses require appeal to perceptual scenarios that produce common knowledge of facts about that scenario in the perceivers, an explanation is needed of how the perceptual base can produce these facts that does not involve common knowledge. In part two, I argue that solving the base problem is possible on an account of the base as a process maintained by the exercise of an object-involving form of joint know-how. Joint perceivers always enjoy a practical form of knowledge about their target and some perceivers can make this knowledge explicit and then have a minimal form of common knowledge about their target. The base problem has therefore a solution. In part three, I show how this solution of the base problem can be applied to analyses of collective intentionality that include a common-knowledge requirement.

## 2 Common Knowledge and the "Base Problem"

The discussion of joint attention is closely tied to that of perceptual common knowledge. Indeed, Schiffer's analysis assumes and Lewis's analysis allows joint attention as what Lewis calls the "base" of common knowledge. In this section I trace the function of joint attention in their respective accounts and show why their treatments give rise to what I call the "base problem".

<sup>2</sup> What Schiffer means by "mutual knowledge" is what today is usually called "common knowledge", where the former designates a finite iteration of knowledge shared by two participants and the latter refers to an infinite such iteration. I shall be concerned with common knowledge in what follows. Thanks go to one of my reviewers for highlighting the distinction.

Schiffer (1972) defines what he calls "mutual knowledge"<sup>2</sup> as follows:

- S and A mutually know that *p* iff
- S knows that *p*.
- A knows that *p*.
- S knows that A knows that *p*.
- A knows that S knows that *p*.
- S knows that A knows that S knows that *p*.
- A knows that S knows that A knows that *p*.
- Etc.

Schiffer illustrates the structure of mutual knowledge with an example in which this knowledge is produced perceptually. S and A are seated at a table with a candle placed between them, and the mutually known proposition is "there is a candle on the table". Schiffer then asks how S and A each know that the other perceiver knows that there is a candle on the table. To answer that question, he introduces the concept of "normalcy". A "normal" person is "a person with normal sense faculties, intelligence, and experience" (31). If such a person "has his eyes open and his head facing an object of a certain size (etc.), then that person will see that an object of a certain sort is before him". Then, if S knows that A is normal and sees that A's eyes are open and his head is facing the candle, S knows that A knows that there is a candle on the table; and so forth. Schiffer says that the resulting regress is "perfectly harmless" and that "the phenomenon which obtains in this case" is general: "it will obtain, broadly speaking, whenever S and A know that *p*, know that each other knows that *p*, and all of the relevant facts are "out in the open" (32).

The "relevant facts" in question must include the condition of normalcy. Suppose S and A are normal in Schiffer's sense but the fact that they are is not out in the open between them. Then S and A are not entitled to the inference that the other person knows that there is a candle on the table, and they are consequently not in a position to know in common that *p*. The epistemic openness of normalcy is a condition of common knowledge.

I now argue that on one possible construal this condition raises a problem for Schiffer's contention that the iterative regress contained in the analysis of common knowledge is harmless. On this construal, the openness of the common knowledge-producing perceptual scenario is not simply a feature of the perceivers' experience but is in need of further analysis. Then the question arises what makes it the case that the normalcy of A and S is out in the open between them. It would not be plausible to argue that it is normal for normal people to know that they meet the conditions of normalcy. This option is unavailable because the openness of normalcy cannot be itself based on an appeal to normalcy if it is to explain how normalcy can be out in the open between people. A different explanation is needed. It is not obvious what shape such an explanation might take. Certainly the normalcy of a situation cannot be known a priori, since

not all people (the neurodivergent, the blind) are normal in Schiffer's sense. Where conditions are known to be normal, at least part of this knowledge must be based on present or past experience. I can come to know that your eyesight is normal, for instance, by observing you competently navigate challenging and novel perceptual environments, or by inferring your normal eyesight from the normalcy of the many other people I've interacted with in the past. I can also come to know that you are of at least average intelligence by observing you solve certain tasks, or again infer it from the intelligence demonstrated by many other people previously. If this is right, knowledge of a person's or context's normalcy can only be inferentially acquired, on the basis of occurrent or past perceptual information or, perhaps, by being told.

This poses a problem for the requirement that the normalcy of a situation in which common knowledge is available has to be out in the open between perceivers. Since even well-justified inferences to normalcy can lead to false conclusions, there is no collective knowing whether the normalcy of a social perceptual scenario is epistemically open in the way required by its role as a condition of common knowledge. Common knowledge turns out to depend on whether agents' justification for their beliefs leads to truth, and there is no knowing whether it does. Even where justification of mutual belief about normalcy leads to truth, this does not entail that knowledge of normalcy is out in the open between perceivers: since competent reasoners know that justification does not always lead to truth, collective agents could never know that they know in common that the condition of normalcy obtains. Yet in that case the introduction of normalcy as a condition of common knowledge makes common knowledge impossible: since the conditions that have to be met for common knowledge to be possible have to be out in the open between agents, since these conditions include normalcy, and since normalcy can never be out in the open, there cannot be common knowledge.

I have just sketched a two-step argument that shows that, on the construal at hand, the appeal to normalcy in Schiffer's analysis of common knowledge makes the argument viciously circular. The first step of the argument is this:

1. Common knowledge requires that the condition of normalcy be met
2. The fact that it is met has to be out in the open between S and A
3. For a fact to be out in the open between S and A, they have to know it in common
4. S and A have to know in common that normalcy obtains if they are to have common knowledge

This first step shows that, on one construal, Schiffer's analysis of common knowledge is circular: common

knowledge requires that perceivers' normalcy be common knowledge between them. You may think, as a reviewer did, that this step alone warrants the conclusion that Schiffer's argument, on the construal at issue, is deficient. Then you can simply omit the following discussion. But you could think that the circle laid out in steps 1–4 is not vicious; that there is a straightforward explanation of how perceivers attain common knowledge of normalcy. I now supply an argument, already sketched in my previous remarks, to the extent that there is no such explanation. It will turn out that since S and A can never know in common that the conditions of normalcy are met, S and A cannot know in common that  $p$ . The second step of the argument shows that the circularity of Schiffer's analysis, on the reading at issue, is in fact problematic.

5. Knowledge of normalcy can only be obtained by way of fallible (but rational) inferences
6. Therefore, S and A can each come to falsely (but rationally) believe that normalcy obtains; and they can each falsely (but rationally) believe that the other knows that normalcy obtains
7. Therefore, S and A can each falsely (but rationally) believe that they know in common that normalcy obtains
8. Whenever S and A falsely believe that they know in common that normalcy obtains, they do not know in common that normalcy obtains
9. Then they do not know in common that  $p$  [from (4)]
10. Since S and A cannot generally distinguish between knowing and falsely believing that normalcy obtains, they do not know whether they know in common that  $p$ .
11. Perceivers who know in common that  $p$  always know that they do.
12. S and A do not know in common that  $p$ .

I take (5 to 7) to be uncontroversial. Normalcy cannot be known a priori, knowledge of normalcy is inferential and therefore fallible, and hence it is possible that candidate collective knowers come to falsely believe that they know in common that normalcy obtains. I can falsely believe that you have "normal sense faculties" and therefore perceptually know that there is a candle on the table even though you are blind and don't have this perceptual knowledge; and I can on this basis come to falsely believe that we know in common that there is a candle on the table [steps (8 and 9)]. But since false belief is not generally<sup>3</sup> distinguishable from

<sup>3</sup> There may be some exceptions here. Perhaps I can hold a superstitious belief that I know to be false but am nevertheless unable to give up. These exceptional cases are not relevant for the present discussion though.

knowledge to the person who holds it, the person is not in the relevant cases in a position to distinguish between a situation in which there is common knowledge of normalcy and a situation in which there isn't; since common knowledge of normalcy is a condition of the common knowledge that  $p$ , perceivers cannot know whether they know in common that  $p$  [step (10)]. This is a problem: since knowers who have common knowledge know that they do,<sup>4</sup> Schiffer's perceivers cannot have it [steps (11 and 12)].<sup>5</sup>

A natural reaction to this diagnosis is to argue that it demands too much of common knowledge (and, by implication, of knowledge in general). Lewis's (1969, p. 56) account of common knowledge analyses it in the weaker terms of "reason to believe":

"Let us say that it is *common knowledge* in a population  $P$  that \_\_\_ if and only if some state of affairs  $A$  holds such that:

- (1) Everyone in  $P$  has reason to believe that  $A$  holds.
- (2)  $A$  indicates to everyone in  $P$  that everyone in  $P$  has reason to believe that  $A$  holds.
- (3)  $A$  indicates to everyone in  $P$  that \_\_\_."

Lewis calls  $A$  the "basis for common knowledge in  $P$  that \_\_\_" and claims that " $A$  provides the members of  $P$  with part of what they need to form expectations of arbitrarily high

<sup>4</sup> I simply take this premise for granted; so do e.g. Sperber and Wilson (1995, pp. 19–20) in their argument against the possibility of common knowledge. There may be room for divergent views here. Discussion is beyond the scope of what's possible in this paper, however.

<sup>5</sup> A reviewer worried that this conclusion may pose a problem for my own recommendation, at the end of Sect. 3.2, that we handle the possibility that individual perceivers cannot always distinguish between common knowledge and mutual false belief by appeal to a version of epistemological disjunctivism. Isn't just the same move available to the sceptic about the present conclusion, who may argue that we can handle the possibility of mistaken beliefs about common knowledge of normalcy disjunctively? The move is unavailable to the sceptic though. The problem laid out above arises because common knowledge of normalcy can only be obtained through individual inferential activity. Since the relevant inferences are inductive and hence fallible, perceivers can never know that they know in common that they are normal even where they are. So there is no "good" case that could be disjunctively juxtaposed to mutual false belief about normalcy. By contrast, there is such a case for perceptual common knowledge: on the view I am developing, joint perceivers individually know that they know in common a perceptual fact whenever they do.

<sup>6</sup> A reviewer pointed out that Lewis is not committed to the view that the basis of common knowledge must be perceptual. I agree. But joint attention surely is one possible basis of common knowledge, and if it is described without tacitly importing theoretical commitments about the nature of perception and the like it must therefore be possible to ask whether Lewis's argument succeeds for cases in which the basis of common knowledge is perceptual. The following discussion applies to these cases only.

order, regarding sequences of members of  $P$ , that \_\_\_." You can think of Schiffer's "candle" scenario as such a basis for common knowledge.<sup>6</sup> Then,  $S$  and  $A$  have reason to believe that the candle scenario holds (that they are each looking at the candle between them, in a way that they also see that the other is looking at the candle and seeing them looking at the candle); the candle scenario indicates to  $A$  and  $S$  that they have reason to believe that the scenario holds; and the candle scenario indicates to  $A$  and  $S$  that there is a candle on the table.

On the face of it, this analysis of common knowledge escapes the objection to Schiffer above. If common knowledge is based on a scenario of joint attention in which subjects have reason to believe about each participant that they are jointly attending to a target, and the joint scenario indicates to them a perception-based fact about the target, then there cannot be a case in which they acquire false beliefs about the other's beliefs about the target on the grounds of false (but rational) inferences from what is visible to them. That they come to know these facts in common is guaranteed by the stipulation, right at the outset, that " $A$  holds".

Whether this account works for scenarios in which the base of common knowledge is perceptual depends entirely on how you think about  $A$ . The question arises how to characterize joint attention. One option is to take it that joint attention is itself analysable in terms of common knowledge.<sup>7</sup> Then the proposal is flatly, and viciously, circular: common knowledge of a perceptual proposition is theorized as being based on joint attention, but joint attention is itself explained in terms of the common knowledge of that proposition. Another option is to take it that joint attention is somehow primitive. Then the charge needs to be avoided that joint attention just gets stipulated into existence to serve as the base of perceptual common knowledge.

The problems that arise for Lewis and Schiffer are thus closely related. For Schiffer, the problem is that the description of the perceptual base of common knowledge requires that certain conditions (those of "normalcy") be out in the open between perceivers, where accounting for this openness itself requires an appeal to common knowledge. For Lewis, the description of the perceptual situation that may serve as the base of (true and rational) mutual belief must appeal to joint attention. If joint attention is defined as a perceptual scenario that produces common knowledge in its participants, then this definition again invokes common knowledge. In both cases, the analysis turns out to be circular. Also in both cases, the circle is vicious, since there is no good (non-circular) explanation of how the common

<sup>7</sup> Though such theories are sometimes (e.g., Battich & Geurts 2021) discussed as alternatives to the "experiential" theory of Campbell (2005), I do not know of anyone who explicitly defends such a theory.

knowledge invoked in the description of the perceptual scenario in and about which common knowledge is attained.

I call this the “base problem” arising for Schiffer’s and Lewis’s analyses:

(BP)For theories of common knowledge that require appeal to a perceptual base, a non-viciously circular and non-viciously regressive account is needed of this base that explains how it can produce common knowledge of a perceptual fact in the perceivers who help constitute the base.

### 3 Joint Attention

How might one address the base problem? The intuitive answer I shall be developing is that the experienced social world itself provides the base that makes common knowledge about it possible. The kinds of environmental, mental, and social facts that are out in the open between agents, so that joint action on objects contained in that environment becomes possible, are directly accessible to the agents in ways that it is a condition of their access that they are shared between these agents. This is nothing more than a common-sense description of what happens when we perceive and act together: the world and its objects are perceptually available to us in a social mode that allows us, often effortlessly, to share relevant facts about them with each other and that thus facilitates joint action. The base of perceptual common knowledge, and of the complex social phenomema that build on it, is the social world that we share.

In the next sections, I develop a view that substantiates these intuitions.<sup>8</sup> Here is a sketch of the core idea. Joint attention can be thought of as a process that is maintained by way of the execution of a minimal form of joint know-how (see also Seemann, under review). For a certain class of joint perceivers (broadly, those capable of linguistic communication), the experience of joint attention is apt to produce in them a minimal kind of common knowledge that is “luminous” in Williamson’s (2000) sense. A mental state is luminous just when its subjects know that they are in that state. If common knowledge is luminous, then knowers who know in common that  $p$  always each know that they enjoy this knowledge. Suppose these knowers are linguistically communicating joint perceivers. These perceivers can always know, when jointly attending to a target, the proposition that expresses the target’s location in social space. Since social space is constituted in social interaction, this knowledge is necessarily of a common kind. Since it is produced intentionally, typically in

linguistic communication, it is luminous. The explanation of how joint attention produces common knowledge avoids circularity because it begins with an account of joint know-how whose description does not invoke common knowledge and explains luminous common knowledge by appeal to the possibility of linguistic expression and communication of some facts that are already practically known to the agents who exercise this joint know-how.

#### 3.1 Joint Attention as a Kind of Joint Know-How

Avoiding the base problem requires an account of joint attention that gets by without mentioning perceivers’ epistemically open intentional states and their interrelation. The account that most obviously delivers on this requirement is Campbell’s (2005, 2011) relational and “object-centered” theory of joint attention. Campbell sees joint attention as an experience that has the perceiver, co-perceiver, and a target object as constituents. This triadic experience is said to be a “primitive phenomenon of consciousness”. There is, on this view, a particular kind of experience available to creatures who are jointly attending to objects with others. The other person “enters the individuation of the experience”, which is thus of an irreducibly different kind than an individual perceiver’s experience of an object. On this experiential view, the base problem does not arise: since the experience is thought of in terms of a *sui generis* triadic perceptual relation, no appeal to common knowledge is required in the metaphysical description of the experience.

Campbell’s broad sketch of the experiential view is not without its problems. Battich and Geurts (2021) raise a variety of difficulties that put pressure on the conception of joint attention as primitive in Campbell’s sense. They point out that this conception of joint attention does not block recursion and that joint attention must be preceded by the recognition of the other person as a co-attender. But for that to be possible, something like Schiffer’s normalcy condition has to be met. As they put it, “the bottom line is that *at least some* knowledge must be involved in any analysis of joint attention” (Battich & Geurts 2021, p. 9). There is also the general worry that since, on any plausible account, perceptual experience is inherently perspectival, it is not obvious how to make metaphysical or phenomenological sense of an experience with two perceivers as constituents. How can there be an experience that presents its target as being perceived from a variety of perspectives? It would seem that the joint experience decomposes into two mereological constituents along the lines of Baron-Cohen (1999), who describes joint attention in terms of mutual “seeing-what-the-other-sees”.

Substantiating the experiential view so that it masters at least some of these challenges is possible if you think of

<sup>8</sup> This view is developed more fully in Seemann (2019).

episodes of joint attention as processes that are maintained by the execution of a practical form of knowledge that the perceivers can only deploy in interaction with each other. The description of these interactions has to be purposive but cannot rely on agents' interrelated representational states, such as their intentions, that would be out in the open between them, since then the base problem arises again.<sup>9</sup> The traditional way to think about motor action without relying on intentions that represent their conditions of satisfaction is to subscribe to a view that has its roots in Merleau-Ponty's (1945/2002) concept of "motor intentionality", such as Dreyfus's (1993/2014) notion of "skillful coping" or Gallagher's (2005) and Hutto's (2008) versions of enactivism about the mind and cognition. For the purposes of this paper, my aim is to develop an account that does not require subscription to this kind of view, though it is compatible with it. To this end, I introduce the technical notion of a "doing". Doings are purposive bodily movements that can be described without appeal to intentional state concepts. Thinking of an agent's contribution to the process that constitutes an instance of joint attention in terms of a doing is sufficient for the theory of joint attention I shall be developing. The question of whether doings should be further characterized in terms of some version of motor intentionality or by appeal to interlocking intentional states that are not out in the open between agents remains, for the purposes of this paper, open.

(DOING) A doing is a proprioceptively experienced bodily movement that involves a perceptually present object and that the moving creature prolongs.

The core consideration is that it is always true that agents who act purposefully prolong what they are doing for as long as they are doing what they are doing, regardless of what their reasons are (if any). Compare the notions of "doing" and of "bodily moving". It is not true of the latter notion that agents prolong their movements as long as they are moving. Suppose a doctor is probing your reflexes by tapping your knee, as a result of which your lower leg moves even if you form the firm intention to keep still. Then you are bodily moving even though you are not prolonging the movement while it is going on. So it is informative to say, about doings

in contrast to other kinds of bodily movements, that agents prolong them while they are going on.

Creatures can prolong what they are doing for many reasons. They may prolong what they are doing because they are enjoying the activity or because they are pursuing some external goal, but they can also keep doing what they are doing if they have no apparent reason for doing so at all (think of the doodles you draw while on the phone with someone). Since everything we do eventually comes to an end, the notion of a doing is temporally indexed: it is only within a certain temporal interval that creatures prolong their doings. You thus can be doing something, in my sense of the term, even though what you are doing will end once you have achieved an external goal, or once you don't find it interesting anymore, or once it is terminated by external factors (you have to do something else; you fall asleep). You still prolong the doing as long as it is going on. Call this the "intrinsic motivation" that is inherent to a doing. Intrinsic motivations are unlike distal intentions in that they could not be entertained outside of the doing. They are also different from what Searle (1983) calls "intentions-in-action" in that the prolongation of the doing does not require meeting internally represented conditions of satisfaction: it is not that creatures intend to prolong their doings and can succeed or fail in doing so. If they are doing something, they are necessarily prolonging what they are doing within the doing's temporal boundaries. All this is compatible with the possibility that the doer entertains intentions, distal or not, and that these intentions play a causal, explanatory or justificatory role in the doing. But the notion of a doing is compatible also with the view that intentions play no role in (some of) the things we do and that there nevertheless is a distinction to be drawn between doings and reflex-like bodily movement.<sup>10</sup>

Doings are always carried out in environments that are perceptually present to the agent, and they "involve" objects or scenes in the environment. "Involvement" here is a technical notion designed to capture doers' purposive engagement with the environment while avoiding having to spell out this engagement by appeal to intentions that represent their conditions of satisfaction. Doings involve objects in the sense that interaction with the object

<sup>9</sup> A reviewer pointed out that a description of joint attention that mentions intentional state concepts need not give rise to the base problem. I think that's right as long as this description does not require that joint perceivers' intentional states are out in the open between them, in which case the base problem plainly arises (see Sect. 2 of this paper for discussion). So there may be room here for an analysis of joint attention that makes reference to perceivers' intentional states but does not require that these be out in the open between perceivers. This motivates the description of perceivers' contributions to episodes of joint attention in terms of "doings" that are non-committal with regard to the notion of (non-representational) motor intentionality.

<sup>10</sup> As a reviewer pointed out, these remarks do not amount to a fully formed account of the technical concept of a doing. Such an account would have to engage with the discussion about motor intentionality and practical knowledge and will have to wait for another occasion. I hope, though, that enough has been said to help substantiate the view of joint attention I shall be developing in what follows. The important thing to bear in mind is, to repeat, that the notion of a "doing" is a technical device whose sole purpose it is to aid in the development of a theory of joint know-how and joint attention while remaining neutral on the question of motor intentionality.

constitutes the doing in question. For example, touching an object constitutively involves the object: you can only touch it if there is in fact direct contact between your body and the object. Touch does thus not have conditions of satisfaction that can be spelled out relative to physical contact (though “trying” or “intending” to touch has such conditions). Other kinds of doings can then be modelled on touch. Suppose you are pointing at a distal object, where the pointing qualifies as a doing (it is being prolonged while it is going on; it makes use of proprioception; it involves the distal object). The pointing involves the object in the sense that the doing could not take place without the object being pointed at. “Involvement” thus does not require physical manipulation. It only requires that specifying the doing, as the kind of doing it is, constitutively includes mentioning the object, so that you could not execute the doing if the object were not there, say. Like knowing and perceiving (but unlike believing), doing is in this sense factive.

Some doings are carried out with other people. And some of these social doings can only be done with other people. You cannot play a game of tennis by yourself, you cannot seesaw by yourself, and you cannot jointly attend to a target by yourself. Call these doings “joint”:

(JD) A joint doing consists of at least two creatures’ bodily movements by way of which they prolong what they are doing, where the doing involves one object that is perceptually present to both creatures and where its prolongation requires that each creature co-ordinate its own movements with the movements of the perceptually present other creature.

More would need to be said about the crucial notion of co-ordination for a complete account, but JD is sufficient for present purposes. I understand joint attention as a kind of joint doing whose objects are, or could be, out of reach and that thus requires participants to deploy techniques, such as perceptual attention, that enable them to involve such objects. The participants in a joint doing exercise a kind of joint know-how: they know how to co-ordinate their movements with those of their co-agent so as to prolong what they are jointly doing. This minimal account of joint know-how, based as it is on the technical notion of a doing, gets by without requiring either subscription to or rejection of (some version of) the notion of motor intentionality. It only requires that the contributions of agents by way of which they maintain an episode of joint attention are describable as purposive movements that are being prolonged by these agents while they are going on. This requirement has two implications for the view of joint attention I am developing.

<sup>11</sup> This consideration also plays an important role in Birch’s (2018) account of joint know-how.

First, on the resulting view joint attention is something agents do purposively. Joint attention is always endogenous; it could not be in its entirety the consequence of external factors. This is plausible: even though you and I can be made to look at the same object by external factors (a loud bang; a flash of light), this does not by itself to joint attention. Joint attention always requires purposive co-ordination of bodily movement. Secondly, participants in joint attention always have an intrinsic motivation to prolong the episode for as long as they do. Even where you and I jointly attend to the wallet I am forcing you to hand over to me at gunpoint, you are intrinsically motivated to attend to the wallet with me while the episode is going on. This is compatible with you not wanting to hand over your wallet to me and not wanting to be interacting with me at all. It only amounts to the difference between you attending to the wallet with me and refusing to coordinate your attention with mine, for instance by directing your gaze elsewhere.

### 3.2 Social Triangulation and Spatial Common Knowledge

Joint agents who prolong an episode of joint attention co-ordinate their movements with those of their co-agents and co-perceivers. This requires them to adapt their movements to those of the other agent and to put the other agent in a position to adapt their movements in turn. For this to be possible, they cannot simply react to the other’s contributions. They have to take an active role to put the other agent in a position to contribute to the joint doing.<sup>11</sup> Consider a case of joint attention: if you and I are to jointly attend to a target, it is not sufficient that I follow your pointing gesture so that my gaze comes to focus on the object you are making salient. My own movements have to contribute to what we are jointly doing. That is, I have to check whether the object I am focusing on really is the one you are making salient and I have to move so as to make that object salient to you. In short, if we are to jointly attend to a target, we have to point it out and keep it salient *to each other*. This requires that we each triangulate the target’s location relative to the standpoint of the other person. Call this, “social triangulation”. Perceivers prolong an episode of joint attention by participating in a continuous process of making and keeping the target salient to each other by socially triangulating its location. When they are involved in this kind of mutual triangulation, they operate in what I call “social space”. Social space is a spatial framework in which targets are mutually singled out relative to the standpoints of agents’ co-perceivers. It thus is a framework in which not only one’s own but also one’s co-agent’s location is presented as a standpoint.

Perceivers who jointly attend to a target exercise a joint form of know-how, as described in the previous section.

They then enjoy a practical kind of spatial knowledge: they always know where the target is located relative to the standpoint of their co-perceiver. This knowledge does not entail that perceivers know where the object is in allocentric space: successful social triangulation is possible even if the target's location in allocentric space is misrepresented (by a mirror arrangement, for instance). Since social triangulation is dynamic (it requires that each perceiver continuously adapt their pointing gestures and other motor movements to those of their co-perceiver), the practical knowledge of a target's location in social space is of a joint kind: no individual agent could have it on their own.<sup>12</sup>

Some joint perceivers are capable of entertaining and linguistically communicating propositions that express the location of the target object in social space. These perceivers can communicate their practical knowledge of a target's location in social space to their co-perceivers by saying things like,

(PK) THIS\* is the location L of the target T we are looking at,

Where their utterance of PK is accompanied by the kind of pointing gesture that is apt to help prolong an episode of joint attention and the uttered token of "THIS\*" refers to the location of the target in the social spatial framework in which the involved perceivers operate. When they communicate with their co-perceivers by uttering PK and accompanying their utterance with the right kind of pointing gesture, speakers and hearers acquire luminous common knowledge of the object's location in social space. It is a form of common knowledge because no speaker could entertain it on their own, because knowing the proposition expressed by an utterance of PK requires that one's addressee know it also and because it is always true that if one of the communicators knows it, then each communicator knows that each communicator knows it.<sup>13</sup> The communication that takes place between the speaker and the hearer of PK can only be an exchange between two participants in an episode of joint attention. It expresses and makes explicit spatial knowledge that speaker and hearers, as joint perceivers and agents, already possess in practical form.

In joint perceptual contexts, PK is always true. The experience of joint attention supplies perceivers with

reasons for PK. When asked why they are saying that PK, a speaker can always reinforce the salience of T by pointing it out to the hearer. Joint scenarios thus meet the conditions stipulated by Lewis: in an episode of joint attention, all joint perceivers have reason to believe that the perceptual scenario is joint; the joint scenario indicates to all perceivers that all perceivers have reason to believe that the perceptual scenario is joint; and the perceptual scenario indicates to all perceivers that PK. Once perceivers know in common that PK, they can construe iterations of what each perceiver knows about each perceiver's knowledge that PK. Thus common knowledge is luminous: when perceivers know in common that PK, they each know that each knows that they know in common that PK. However, its luminosity is epistemologically unproblematic. The iterations that make PK luminous are entailed by the common knowledge that PK; they do not constitute this knowledge. The regress really is harmless.

This account avoids the base problem as it arises for Lewis and Schiffer. The perceptual base of common knowledge is joint attention, joint attention is defined as a process that is maintained by way of an exercise in joint know-how involving a distal object, and joint know-how is describable without appeal to agents' cognitive (or otherwise intentional) states. Some of these agents can express their joint know-how linguistically. When these agents manage to linguistically communicate to each other the location of the target of their joint attention in social space, they know this location in common and this common knowledge is necessarily luminous. This account of joint attention can be inserted into Lewis's account so that it meets the conditions of the base of common knowledge. For Schiffer, the problem was that on one reading the condition of the openness of normalcy could only be met if it was commonly known between perceivers, which led to a vicious circle. The problem can be avoided on an account of joint attention on which the involved perceivers' normalcy is out in the open between perceivers in virtue of their experience of jointly attending to a target. The enacted theory of joint attention delivers such an account. The experience of joint attention is the experience of a process that is maintained through the exercise of a joint know-how. When a perceiver enjoys this experience, she is participating in a joint doing. When such perceivers can linguistically communicate the location of the target of their joint attention to each other, they luminously know the location of this target in common. These perceivers then also are in a position to know that the conditions of normalcy obtain, and that this is common knowledge between them. The enacted account turns things upside down with regard to the condition of normalcy: it suggests that where perceivers exercise a joint form of know-how that involves distal perceptual objects, they are normal in the required sense. Common knowledge of normalcy

<sup>12</sup> See Seemann (under review) for a detailed argument.

<sup>13</sup> The perceiver who understands a co-perceiver's utterance of PK comes to know that the referent of the speaker's utterance of a token of THIS\* (accompanied by a suitable pointing gesture) is the object of the speaker's referential intention. As soon as the hearer has recognized the speaker's intention to make salient the referent of her utterance to him, speaker and hearer know in common which object (as specified by its location) they are communicating about. The assertion and understanding of PK that constitute communication occur on the personal level and therefore the knowledge produced by the communication is luminous. See Seemann (2019) for exposition.



thus conceived is a consequence, not a prerequisite, of successfully exercised joint know how.

The view that joint attention is a process that is maintained by perceivers' joint know-how is attractive not only because it helps avoid the base problem for some classic theories of common knowledge. It also provides answers to some of the questions arising for the experiential account of joint attention proposed by Campbell and others. The enacted account conceives joint attention as a temporally extended process that is prolonged by way of the contributions of its participants. Each participant's experience then is of the process that they help constitute. On this view, the other person enters the perceiver's experience because the experience is of a process that is co-constituted by the contributions of each perceiver. And the experience has three constituents because it presents the target as singled out via a process of triangulation that, for each perceiver, takes the co-perceiver's location as a standpoint. All this is compatible with the core tenet of the experiential view that the experience of joint attention is primitive. It is primitive in the sense that it cannot be reductively analysed in terms of the cognitive or phenomenal states of its participants. Of course, it is not primitive in the sense that nothing more could be said about it, but this is not a demand that could be met by any plausible theory.

For all this, the sceptic can still respond that the enacted account of joint attention does not explain how to handle the possibility that an individual perceiver may be mistaken about participating in a joint scenario. You can come to falsely believe that I am jointly attending to a target with you by misconstruing my direction of gaze, for instance. It is even possible that two perceivers each come to falsely believe that PK and thus each falsely believe that they know in common that PK. The objection is that common knowledge, though luminous, does not forestall the possibility of mutual false belief. This is, of course, true. But the argument is misguided. The challenge gets off the ground because of an implicit commitment to reductionism about the experience of joint attention. It takes it, tacitly, that what you can think of as the "subject base" of the experience is the individual perceiver, and that, therefore, the possibility of falsidical experience needs to be settled at the individual level. On such a view, however, joint attention could not produce luminous common knowledge and the base problem would have no solution. The enacted view is designed precisely to avoid this unattractive conclusion. The defender of enactivism does not have to deny that the bearers of joint experiences are individuals. But since the experience is of a joint process that resists decomposition into the individual contributions of its participants and that produces a minimal form of common knowledge in these participants, the experience of the participant is of an epistemologically different kind than that of the solo perceiver who falsely

believes that she is participating in such a process. The defender of the enacted view is thus committed to a social version of epistemological disjunctivism about experience (Seemann 2019, pp. 67–72). On this view, the observation that perceptual mistakes are always possible does not imply that joint experiences are to be individuated by appeal to perceivers' individual beliefs about the character of their experience.

#### 4 The Base Problem and Collective Intentionality

I have argued that some classic analyses of common knowledge face the base problem and I have suggested that an enacted theory of joint attention can help avoid this problem. In this section I show that this theory does not just remove a difficulty for these analyses of common knowledge. It is also important in the context of some discussions about joint action and collective intentionality. The relevant theories are sometimes called "reductive" (Blomberg 2015). They are given this label because they attempt to explain how agents can act jointly via an analysis of the interrelation of individuals' mental states. Broadly, for joint action to be possible, each agent has to intend to jointly do  $x$  with the other person; each agent has to intend to contribute what is required of them to bring about  $x$ ; each agent believes that the other person intends to contribute what is required of them to bring about  $x$ ; and they each to contribute what is required of them because they believe that others intend to contribute what is required of them also.<sup>14</sup> So joint action is explained reductively in terms of a complex structure of individuals' intentions that interlock in certain ways.

Theories of collective intentionality that are reductive in this sense are not, however, eliminative. They do not propose that a complete explanation of complex social phenomena is possible without appeal to collectivity concepts of any kind. A popular move is to require that the intentions and their interrelations that explain how agents can act jointly on shared goals are "out in the open" between them.<sup>15</sup> To explain this cognitive openness of the situation in which collective intentions can be formed and executed, some theories introduce a common knowledge condition without which the reductive analysis remains incomplete. The reductivism of this family of approaches to collective

<sup>14</sup> This rendering of how individuals' intentions have to interrelate if they are to constitute a collective intention is loosely adapted from Pettit and Schweikard (2006).

<sup>15</sup> E.g., Bratman (1992); Cohen and Levesque (1991); Tollefsen (2005); Tuomela and Miller (1988). For a critical analysis of this openness requirement, see Blomberg (2015).

intentionality requires the epistemic openness of the relation between their relevant intentions and subplans and thus avoids eliminativism about the social by stipulating a common knowledge condition.

If the argument in the first part of this paper is correct and classic accounts of common knowledge face the base problem, then the problem arises also for reductive theories of collective intentionality insofar as they rely on these or related accounts. Consider Bratman (1999, p. 121):

We intend to *J* if and only if

1. (a) I intend that we *J* and (b) you intend that we *J*.
2. I intend that we *J* in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b; you intend that we *J* in accordance with and because of 1a, 1b, and meshing subplans of 1a and 1b.
3. 1 and 2 are common knowledge between us.

(3) Introduces the common knowledge condition that makes the theory susceptible to the base problem. Sometimes the condition can be satisfied by linguistic communication alone, in the absence of a perceptual context in which the action is executed. Suppose we intend (Searle's example) to make a sauce hollandaise together later, and we now linguistically communicate our respective intentions and subplans (I will later pour the ingredients and you will mix them while I pour) to each other. Then we have satisfied the common knowledge condition. There is a question whether this kind of linguistic communication itself requires some kind of joint perceptual context, but I am leaving this question aside for present purposes. I am interested in cases in which a collective intention is directly tied to a perceptual joint action context in which agents coordinate their motor movements in pursuit of a shared goal. Suppose we are in our kitchen and the communication of our intentions and subplans is at least in parts effected by our actions: having established earlier that we want to make a sauce together but having left open how, I now fetch the ingredients, demonstratively put the bowl on the table and hand you the mixer; you take the mixer from me; I begin to pour the ingredients and you mix them. Then some of the epistemic openness that satisfies the common knowledge condition is perceptually constituted. In these cases, going by BP, a non-viciously circular and non-viciously regressive account is needed of the perceptual base that makes available common knowledge of the agents' intentions and thus ensures that these intentions are out in the open between perceivers. I have offered such an account: the enacted view of joint attention explains how joint perceivers capable of linguistically expressing some of the practical knowledge by way of whose execution they prolong a process of joint attention can acquire a minimal

form of perceptual common knowledge about the location of its target. Elsewhere (Seemann 2019, pp. 73–77) I have called this kind of common knowledge “primary”: it could only be enjoyed in virtue of agents' exercise of a perception-based form of joint know-how, and it is always luminously enjoyed by joint agents who can linguistically communicate to each other propositions of the form PK.

Once this primary kind of perceptual common knowledge is established through linguistic communication (and the performance of suitable accompanying demonstrative gestures), speakers can expand on it. They can, for instance, communicate to each other their intentions about the referent of a token of “THIS\*”. In this way they can acquire a secondary kind of perception-based common knowledge. In the above example, our meshing subplans in the making of the sauce are practically known by us in the way we prolong what we are jointly doing, and they are thus out in the open even if they are not explicit common knowledge between us. But we can communicate to each other propositions that express the subplans by way of whose execution we seek to realise our collective intention, in ways that demonstratively reference what we are doing. I can assert, while moving so as to contribute to our joint making of the sauce, that I intend to contribute to my intention to make a sauce hollandaise with you by pouring the ingredients by way of my present movements and you can assert that you intend to contribute to your intention to make a sauce hollandaise with me by mixing the ingredients by way of your present movements. In the kinds of projects under consideration, what we are each doing requires coordination between our respective contributions. We can then formulate propositions that connect our collective intentions to the joint doings by way of which we realize these intentions. They take something like the following form:

(CP) I intend that we *J* by me  $\varphi$ -ing like \*thus and you  $\Psi$ -ing like \*so,

where “*J*” stands for our joint action, “ $\varphi$ ” and “ $\Psi$ ” for the subplans by way of which we realise *J*, and “\*thus” and “\*so” are the demonstratives whose token utterance refers to the motor movements by way of which we carry out our subplans. In the kinds of scenarios in which an utterance of “\*thus” or “\*so” refers to the motor movements by way of which agents participate in a joint doing, these subplans necessarily mesh in Bratman's sense. When participants in a joint doing communicate with each other by expressing propositions of the form CP and accompanying them with suitable demonstrative gestures, they acquire propositional common knowledge of what it is that they are jointly doing. CP then plays exactly the same role for the intentions with which joint agents act in the pursuit of collective goals as PK does for the location of the object of a joint doing: in

each case, the communication of these propositions between the participants in a joint doing transforms joint know-how into propositional common knowledge. Agents who can communicate and thus come to know CP can always also communicate and thus come to know in common that PK: in order to know what subplans agents are jointly executing in pursuit of a joint intention, they have to know in common where the target is that is involved in the execution of their subplans. The converse, however, is not true: it is always possible that speakers who seek to expand their collective knowledge by communicating to each other propositions expressing what they are doing fail to do so. Agents can be joint doers, know in common where the object is that is involved in what they are doing, and yet be mistaken about the collective intentions with which they believe to be executing their actions. Secondary common knowledge presupposes primary common knowledge, but the reverse is not true.

I have argued that on the enacted view of joint attention, the base problem can be avoided for reductive analyses of collective intentions that rely on classic theories of common knowledge, at least for scenarios in which there is a direct demonstrative connection between agents' joint doings and the intentions with which they are acting. The general idea is the same as in the discussion of common knowledge in part two of this paper: joint agents exercise an object-involving form of know-how by which they prolong what they are doing, and this know-how is describable without appeal to interlocking intentional states that are out in the open between the agents. They can then, given the requisite conceptual and linguistic capacities, communicate to each other propositions that express facts about the target they are acting on or the joint doing they are performing, and they can demonstratively connect these propositions to the doing and its target. This strategy always yields a minimal primary kind of spatial common knowledge that, because it is effected through linguistic communication, is always luminous and therefore solves the base problem for common knowledge. But it may also produce common knowledge of the subplans by way of whose joint execution agents pursue their collective intentions. Differently from the primary kind, this secondary form of common knowledge is not necessarily available to linguistic joint agents: they may be involved in a joint doing but form mistaken beliefs, even false mutual beliefs, about the intentions with which they are acting.

**Data availability** n/a

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless

indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Baron-Cohen S (1999) *Mindblindness*. MIT Press, Cambridge MA
- Battich L, Geurts B (2021) Joint attention and perceptual experience. *Synthese* 198(9):8809–8822
- Birch J (2018) Joint know-how. *Philos Stud* 176:3329–3352
- Blomberg O (2015) Common knowledge and reductionism about shared agency. *Australas J Philos* 94(2):315–326
- Bratman M (1992) Shared cooperative activity. *Philos Rev* 101:327–341
- Bratman M (1999) *Faces of intention: selected essays on intention and agency*. Cambridge University Press, Cambridge
- Campbell J (2005) Joint attention and common knowledge. In: Eilan N, Hoerl C, McCormack T, Roessler J (eds) *Joint attention: communication and other minds*. Oxford University Press, Oxford, pp 287–297
- Campbell J (2011) An object-dependent perspective on joint attention. In: Seemann A (ed) *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience*. MIT Press, Cambridge MA, pp 415–430
- Cohen P, Levesque H (1991) Teamwork. *Noûs* 25(4):487–512
- Dreyfus H (1993/2014) Heidegger's critique of the Husserl/Searle account of intentionality. In *Skillful coping: essays on the phenomenology of everyday perception and action*. Oxford University Press, Oxford, pp 76–91
- Eilan N, Hoerl C, McCormack T, Roessler J (2005) *Joint attention: communication and other minds*. Oxford University Press, Oxford
- Gallagher S (2005) *How the body shapes the mind*. Oxford University Press, Oxford
- Hutto D (2008) *Folk psychological narratives: the sociocultural basis of understanding reasons*. MIT Press, Cambridge MA
- Lewis D (1969) *Convention*. Harvard University Press, Cambridge MA
- Merleau-Ponty M (1945/2002) *Phenomenology of Perception*. Routledge, Milton Park
- Moll H, Kadipasaoglu D (2013) The primacy of social over visual perspective-taking. *Front Hum Neurosci*. <https://doi.org/10.3389/fnhum.2013.00558>
- Moll H, Meltzoff A (2011) Joint attention as the fundamental basis of taking perspectives. In: Seemann A (ed) *Joint attention: new developments in psychology, philosophy of mind, and social neuroscience*. MIT Press, Cambridge MA, pp 393–413
- Pettit P, Schweikard D (2006) Joint actions and group agents. *Philos Soc Sci* 36(1):18–39
- Schiffer S (1972) *Meaning*. Oxford University Press, Oxford
- Searle J (1983) *Intentionality: an essay in the philosophy of mind*. Cambridge University Press, Cambridge
- Seemann A (2019) *The shared world: perceptual common knowledge, demonstrative communication, and social space*. MIT Press, Cambridge MA
- Seemann A (2021) An externalist theory of social understanding: interaction, psychological models, and the frame problem. *Rev Philos Psychol*. <https://doi.org/10.1007/s13164-021-00584-z>
- Seemann A (2022) The public character of visual objects: shape perception, joint attention, and standpoint transcendence. *Phenomenol Cognit Sci*. <https://doi.org/10.1007/s11097-022-09842-6>
- Seemann A (under review) Joint attention and joint know-how.

- Sperber D, Wilson D (1995) *Relevance: communication & cognition*. Blackwell, Oxford
- Stanley J, Williamson T (2017) Skill. *Noûs* 51(4):713–726
- Tollefsen D (2005) Let's pretend! Children and joint action. *Philos Soc Sci* 35(1):75–97
- Tomasello M (2014) Joint attention as social cognition. In: Moore C, Dunham PJ (eds) *Joint attention: its origin and role in development*. Psychology Press, New York
- Tuomela R, Miller K (1988) We-intentions. *Philos Stud* 53(3):367–389
- Williamson T (2000) *Knowledge and its limits*. Oxford University Press, Oxford
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.