

## Joint Attention: Its Nature, Reflexivity, and Relation to Common Knowledge

---

*Christopher Peacocke*

Two parents are watching their son take his first upright steps in learning to walk. Here we have a paradigm of joint attention. The two parents are attending to their son; they are aware of each other's attention to their son; and all this attention is wholly overt. Everything is in the open, nothing is hidden. In what does this openness consist? Can we characterize it explicitly, without using metaphors?

The parents are likely to switch their attention to each other and smile during this episode. This would be a case of what is sometimes called 'contact attention'; and here too everything is in the open. Again, we can ask: what is a literal, explicit characterization of this openness?

As a stipulation, and following some but not all writers in the field, I will restrict the terms 'joint attention' and 'contact attention' to episodes which fully possess this openness. So our first task is to say what, constitutively, joint attention and contact attention are. We cannot hope to be clear about the relations of joint attention and contact attention to other phenomena and capacities until we have an answer to this foundational question. The question is not one of giving an analysis of the meaning of some word in English which has a determinate sense which is hard to make explicit. We use such metaphorical terms as 'openness' precisely because there is no such English word. So our task here is not conceptual analysis: it is rather that of characterizing properly a complex psychological phenomenon, of saying what it is.

After offering such a characterization, I will go on to consider the relation of joint attention to common or mutual knowledge. I will argue that the openness of joint attention is more fundamental than mutual knowledge; that in many cases it makes mutual knowledge possible; and that a range of phenomena that have

I thank Naomi Eilan for encouraging me to write up in more systematic form the remarks I made about the relations between joint attention and common knowledge in the general discussion session at the 1999 Warwick Conference on Joint Attention. (This is also known as calling my bluff.) I have been greatly helped by the challenge of responding to Stephen Schiffer's incisive critical comments on an earlier draft; by Gilbert Harman's guidance on the massive existing literature on some of the topics discussed here; by the editorial comments of Christoph Hoerl, Naomi Eilan, and Johannes Roessler; and by the discussion at a presentation of this material at the NEH Summer Institute on Consciousness and Intentionality at UCSC in 2002, run by David Chalmers and David Hoy, where I learned from the comments of George Downing, Michelle Montague, Scott Sturgeon, and Deborah Tollefsen.

been characterized in terms of mutual knowledge should rather be elucidated in terms of joint attention. If I am right, joint attention and its properties should receive a much more prominent position in the philosophy of mind and language than it has been accorded hitherto.

### 1. CHARACTERIZING THE OPENNESS OF JOINT ATTENTION

Joint attention involves much more than two subjects attending to the same object. Two subjects could each be attending to the same object without either being aware that the other is attending to the same object. More strikingly, two subjects could each be aware, and it could also be part of the content of their experience, that the other is attending to the same object, without the episode having the openness of joint attention and contact attention. Consider two people who are standing facing each other, separated by a thick pane of glass. Suppose each person falsely believes that this glass is a one-way mirror, allowing him to see the other, but preventing the other from seeing him. So each really sees the other, while believing the other cannot see him. This is far from having the openness of contact attention. Similarly, we can suppose that in this situation, both are attending to something—an animal, say—in their common field of view, off to one side of the glass between them. Each may have a genuine perception of the other attending to exactly the same thing as he is attending to, viz. the animal. But because each believes that the other cannot see him, this too is far from having the openness present in our paradigm cases of joint attention.

It might be suggested that these are not cases of joint attention, because each person does not perceive that the other is jointly attending with him in our strong, favoured sense of joint attention. That is a true statement about these cases. It is indeed true that in joint attention, subjects do perceive, or can be aware, that they are jointly attending. But this point cannot answer our original question, for it embeds the notion of joint attention within the content of psychological states such as attention, perception, and awareness. If our task is to say what joint attention and contact attention are, we do not fully answer that question by giving a condition which embeds that very notion within certain mental states.

The point bears upon the suggestion that what is distinctive of joint attention is that the co-attender figures, as someone who is co-attending to the same object, as a constituent of the subject's perceptual experience.<sup>1</sup> What does 'co-attender' mean in such a condition? If 'co-attender' means just 'attends to the same object', that condition will be met in our example of the glass barrier that is falsely believed to be a one-way mirror, in the case in which both subjects attend to the

<sup>1</sup> This proposal is made by John Campbell in Chapter 13 above, and in Campbell, 2002, ch. 8.

animal off to one side. In that example, each perceives that the other is co-attending in that sense. If 'co-attender' means something stronger, and implies full joint attention to the object to which both are attending, the notion of a co-attender simply embeds the property which is to be explained, the openness of joint attention.<sup>2</sup> This is not to say that such embeddings do not provide an important constraint upon what joint attention is. It is just to say that they cannot be the full account. Compare: to be fashionable, something must be believed to be fashionable. But this cannot be a complete account of what it is to be fashionable. We cannot fully or completely individuate the property of being fashionable by saying that it is that property *P* such that to have *P*, something must be believed to be *P*. Too many other properties besides that of being fashionable meet this condition; and correspondingly, it intuitively leaves out too much of what is involved in being fashionable.

How, then, are we to say, in a more informative way, what it is that is missing from the glass-barrier example, and is present in paradigm cases of genuine joint attention?

At this point, it is almost a reflex of a contemporary philosopher of mind to suggest that what is missing from the glass-barrier example is arbitrarily high iteration of some mental state. This would be an instance of the general style of approach that has proved so fruitful in the case of the special kind of overttness found in the much-discussed phenomenon of mutual or common knowledge, so carefully identified and discussed by David Lewis (1969, ch. 2, sect. 1) and Stephen Schiffer (1988, ch. 2, sect. 2). Let us follow Schiffer and take this as our definition of *x* and *y*'s having mutual knowledge\* that *p* (the asterisk, is simply to indicate that we have a defined term here):

*x* knows that *p*  
*y* knows that *p*  
*x* knows that *y* knows that *p*  
*y* knows that *x* knows that *p*  
*x* knows that *y* knows that *x* knows that *p*  
*y* knows that *x* knows that *y* knows that *p*  
 etc.

If we have only a finite initial segment of this series of conditions holding of *x* and *y*, we do not have fully overt knowledge between *x* and *y* that *p*. It has seemed plausible to suppose that in at least some cases, the overttness of *x* and *y*'s knowledge consists in the holding of every proposition in this iteration. Could we not simply apply the same kind of idea to the case of the overttness of full joint attention?

<sup>2</sup> This means that if co-attention is construed in the weaker, first way, there will be a gap in Campbell's argument that perception of co-attention provides a basis for mutual knowledge.

There is a problem in trying to do so, and the problem stems from one of the differences between knowledge and perception. It is part of any plausible defence of the view that Schiffer's definition actually applies to a particular pair of individuals that for a person to have a belief captured in one of the longer iterations of belief, it is required only that the believed content is something the person could infer from what he indisputably currently believes. The thought of the believed content in a complex embedding does not have to enter his consciousness for the attribution of the belief to be correct. Many beliefs are indeed like this, as Schiffer notes: 'I trust that it is true of each philosophy don in Oxford that he knows that his maternal grandmother was never married to Benito Mussolini' (1988, p. 36). Now suppose we offered this hierarchy as distinctive of what is involved in joint attention of  $x$  and  $y$  to  $o$ :

- $x$  perceives that  $x$  and  $y$  are attending to  $o$
- $y$  perceives that  $x$  and  $y$  are attending to  $o$
- $x$  perceives that  $y$  perceives that  $x$  and  $y$  are attending to  $o$
- $y$  perceives that  $x$  perceives that  $x$  and  $y$  are attending to  $o$
- $x$  perceives that  $y$  perceives that  $x$  perceives that  $x$  and  $y$  are attending to  $o$
- etc.

The problem is that the observation which defends the applicability to real people of the definition of mutual knowledge\* does not carry over to the perceptual case, nor to any other state of occurrent awareness. The sense in which each philosophy don in Oxford has the belief about his maternal grandmother is a counterfactual sense. It is something he would infer from currently stored beliefs by principles he already accepts. But perception and other states of occurrent awareness are not merely counterfactual at all. Someone perceives something to be the case only if in the actual world he is in a conscious state with the representational content of what he perceives to be the case. And it is quite implausible that in all cases that display the openness of joint attention, subjects are in the perceptual states mentioned above, with arbitrarily complex embeddings of the 'perceives that' operation, or any other operator expressing occurrent awareness.

The situation is yet more demanding on the philosophical theorist who wants to give a constitutive account of this openness. Even though arbitrarily high iterations of 'perceive that' or 'is occurrently aware that' do not obtain, nevertheless the openness of genuine joint attention is not something dispositional or counterfactual either. In fact, intuitively, what we are still characterizing metaphorically as the openness of the situation of joint attention and of contact attention is not merely something which exists: it also seems to be present to the consciousness of the participants. How are we to account for this combination of characteristics, the presence of such openness in consciousness, combined with the absence of arbitrarily complex embeddings within 'perceives that...?'

We can make a start on this task by characterizing what it is for a state of affairs to have what I will label 'mutual open-ended perceptual availability' to two subjects. To be thus available, a state of affairs must meet the following condition:

If the obtaining of the state of affairs, and the operation of perceptual and attentional mechanisms in the two subjects, bring it about that one of them perceives that the state of affairs obtains, or bring it about that one of them perceives that the other perceives that it does, or brings it about . . . etc., then the state of affairs (thus brought about) of his so perceiving is available for the other to perceive.

The other does not actually have to perceive the state of affairs so brought about for mutual open-ended availability to obtain. The availability to the other of the state of affairs brought about suffices. For the perceptual state of one person to be available to a second is for there to be information in the public environment of the second person which permits a sound computation to the content that the original person is in the perceptual state, using computational principles of the same kind (possibly differing only in complexity of the contents involved) as those already employed by the second person. Information can be available to a subject, but not in fact be accessed by that subject. This is not the same as its being hidden. In the case of joint attention, the state of affairs of the two subjects each attending to a given object is one which has mutual open-ended perceptual availability to them.

This description of open-ended availability is a little too stringent. You may see the other as seeing that you see that *p*. Now suppose the other then becomes occurrently aware of this third-order seeing of yours. Do we really want a characterization of open-endedness that excludes iterations that, at the higher levels, involve occurrent awareness that is not actually perceptual? That seems too strong. The distinctive openness can be present provided occurrent awarenesses at higher levels are available, even if they are not actually perceptual (they remain different from the results of conscious inference). We should indeed still require genuine perception that the other perceives that *p*. It is distinctive of the phenomenon we seek to characterize that any iterations are based on conscious awareness, rather than reached by personal-level inference. We can relax the definition of mutual open-ended perceptual availability to accommodate these points. Henceforth I take the mutual open-ended availability of a state of affairs *s* to two people to mean this:

Each perceives that the other perceives that *s* obtains; and if either is occurrently aware that the other is aware that he is aware . . . that *s* obtains, then the state of affairs of his being so occurrently aware is available to the other's occurrent awareness.

Here I take it that perceiving that something is so is one, but not the only, form of occurrent awareness that it is so.

No finite mind can accommodate arbitrarily long embeddings of contents beginning *perceives that* or *is occurrently aware that* in its representational contents. What is available for a subject to perceive will eventually not be capable of entering the content of the perceiver's states, because of cognitive and computational limitations in the perceiver's psychological economy. These limitations should not be taken as limitations on the openness of the situation of full joint attention, but as limitations on the ability of the subject to take advantage of what is made available by such openness. If we were to have a definition of openness which can be satisfied only by infinite minds, it could not capture the openness which is enjoyed in joint attention by actual finite human beings such as ourselves.

The mutual open-ended perceptual availability of  $x$ 's and  $y$ 's attending to a particular object seems to be a necessary condition for the openness of their joint attention to the object. It is not sufficient. Consider someone who is surprised and delighted when he discovers that when he is attending to an object, the other person notices that fact. This can be the situation of someone who is just coming to appreciate the possibility of joint attention, but is not yet fully engaging in it. This is a case of *incipient* joint attention. Yet this person and his partner's attention to an object may have the property of mutual open-ended perceptual availability. Having this property is consistent with merely incipient joint attention. It is one thing for a state of affairs to have that property; another for the participants to have some awareness that it does. I think this awareness that the case is one of open-ended availability is essential to the full openness of joint attention.

More generally, when there is full joint awareness between two subjects, there is awareness of full joint awareness. That is, full joint awareness has what logicians would call a fixed-point character. The phenomenon of full joint awareness shows that this logical property corresponds to something psychologically real. Now, if mutual open-ended perceptual availability can be present without awareness of its presence, and if full joint attention cannot be present without awareness of *its* presence, then it follows that full joint awareness does not consist merely in mutual open-ended perceptual availability. Mutual open-ended perceptual availability can be at most part of what is involved in full joint attention.

Mutual open-ended perceptual availability is a notion that it takes some intellectual reflection to formulate. It might be doubted for this reason whether it can really enter the content of perceptual experience. I would contest the objection. There are other clear cases in which a person experiences something as being a certain way, does not necessarily have a word or recognitional concept for that way, and in which his experiencing it as being that way consists in his experiencing it as having other features, which he may find it quite hard to articulate. To take a familiar example, someone may experience something as diamond-shaped. Someone can have this experience without having formed a recognitional concept *diamond-shaped*. To experience something as diamond-shaped is to experience it as a shape which is symmetrical about the bisectors of its angles (see Peacocke,

1992, ch. 3). Even quite sophisticated thinkers find it hard to say what it is about something that is diamond-shaped that makes them experience it as so. All the same, the experience of symmetry about the bisectors of the angles is real and effective in producing the distinctive way the whole shape is experienced as being. At a much greater level of complexity, there is no difficulty of principle in two people experiencing their joint attention to an object as having a certain feature, a feature which consists in their experiencing certain properties and relations as instantiated, properties and relations they may find it hard to articulate despite the familiarity of the experience.

We might try another hypothesis, and appeal not just to mutual open-ended perceptual availability, but to awareness thereof. Is awareness by two people, attending to the same object, of the mutual open-ended perceptual availability of their situation sufficient for the intuitive openness which is characteristic of joint attention? Such awareness can be occurrent; and since it does not involve an arbitrarily high iteration of occurrent attitudes, it is immune to the objections against the earlier naïve extension from the case of mutual knowledge. None the less, even considering the possibilities in the abstract, we would have to say that this awareness is still not sufficient for full openness. For someone could be aware that his situation has the property of mutual open-ended perceptual availability, without being aware that the other had the same awareness. One could think of cases in which someone is told that the other, unlike oneself, does not have such awareness, and in which one has come to internalize this, so that it affects one's perceptions of the other.

In actual cases of joint attention, each is aware of the mutual open-ended perceptual availability because he sees the other, he sees the object to which both are attending, and sees the spatial relations between these three things, and the other's perceptual organs. So does it suffice to add a requirement to that effect? We might try to add to the conditions that

$x$  and  $y$  are fully jointly attending to  $o$  iff

(a)  $x$  and  $y$  are attending to  $o$ ;

(b)  $x$  and  $y$  are each aware that their attention in (a) has mutual open-ended perceptual availability;

this further requirement:

the awareness in (b) results from  $x$ 's and  $y$ 's perception of their spatial relations to  $o$ , and to one another, and to the sense-organs involved in their attention in (a)?

That would not suffice. If a person who informs each of  $x$  and  $y$  that their attention in (a) has mutual open-ended perceptual availability does so only when he has checked on the spatial relations of  $x$ ,  $y$ , and  $o$ , and the properties of their sense-organs, the 'results from' relation in the proposed third requirement could hold. But this would not be case of full joint attention. Besides being wrong in detail,

such a style of modification is taking a wrong turn for several reasons. The most fundamental is that we are aiming to characterize the experience of openness itself, not what makes it available.

I suggest that one distinctive feature of full joint attention is as follows. Suppose you are a participant in a situation of full joint awareness. Concerning the total awareness which is involved in your joint attention, you are aware of the following: that both you and the other person are aware that this total awareness exists.

On this view, the total awareness has an indexical intentional content which makes reference to the total awareness itself. Is this coherent and intelligible, and if so, what does it involve?

There are many examples of mental states and events whose intentional contents contain indexical components that make reference to those very same mental states and events. Suppose I am being lazy, and my friend encourages me on urgent practical grounds to start thinking. I may then think:

This thinking is coming too late.

My thought is about the very event of thinking itself; and the thought may very well be true. If I am interrupted in my thinking, that may be my only thought on this occasion. The episode of thinking 'This thinking is coming too late' may be the only thinking to which the demonstrative component 'This thinking' refers. Mental events and states, like linguistic items, can sometimes refer to themselves indexically.

The phenomenon is to be distinguished from cases falling under two other descriptions, cases whose coherence is doubtful. It is *prima facie* plausible that awareness of something is always distinct from what it is awareness of. (A closely related claim would be that awareness that something is the case is always distinct from the holding of the proposition of which it is awareness.) If that is so, no state can be, or consist in part in, an awareness of itself. But the indexical mental self-reference of which I have been speaking does not involve such incoherence. To say that a mental state of awareness has an intentional content which refers to that awareness is not to say that it is an awareness that consists in part in an awareness of itself. The indexical component *this awareness* can refer to an awareness without the awareness referred to having itself as an individuating component. Such reference no more involves a regress of individuation than does the intentional component *this thinking* in 'This thinking is coming too late'. Such a self-referential thought does not require that some event of thinking be a constituent of itself. For something to have an intentional content which refers to itself is not for it to be individuated by its relations to itself. There is no kind of objectionable ungroundedness here.

Indexical mental self-reference is also distinct from the use of such vacuous contents as 'This intentional content is thus-and-so'. Here the indexical component purports to refer to an intentional content; and arguably it cannot succeed



in doing so. The indexical intentional contents with which I am concerned refer, by contrast, not to other intentional contents or their constituents, but to mental events or states. So they do not have this second kind of ungroundedness either.

Indexical self-reference can generate iterations. The extent to which this possibility is realized depends upon what notions feature in the content in which the indexical self-reference is made, and on what principles hold for those notions. To illustrate how indexical self-reference can generate iterations, I will first consider a linguistic case, since it shows the phenomenon in relatively pure form, free of extraneous complications. The basic structures illustrated in this simple case are present also in much more complex examples.

Suppose we have a rather powerful encyclopaedia *E*. It contains not only first-order information about the world, but also contains information about encyclopaedias, including itself. In particular,

(i) The encyclopaedia *E* says:

(C) Rome is Italy's capital and this encyclopaedia *E* contains this conjunction.

We can raise the question of which propositions this encyclopaedia is committed to. There is a notion of commitment, and an entailment-like notion of implication for which we have the following principle of closure of commitment under implication:

(Closure Principle) If *E* is committed to *A*, and *A* implies *B*, then *E* is committed to *B*.

From *E*'s containing the self-referential information (C) together with the Closure Principle, we can obtain arbitrarily complex iterations of 'E is committed to E is committed to ... C', as follows.

From the statement of the example, proposition (i), we have

(ii) *E* is committed to *C*.

Now *C* itself implies that *E* contains *C*, and hence that *E* is committed to *C*. That is, we have

(iii) *C* implies *E* is committed to *C*.

We can now apply the Closure Principle, taking the proposition *C* as the value of 'A' and the proposition 'E is committed to C' as the value of 'B' in our formulation of the Closure Principle. From that Principle, (ii), and (iii), it follows that

(iv) *E* is committed to: *E* is committed to *C*.

That gives our first iteration. We have also just shown:

(v) (ii) implies (iv).

We can now apply the Closure Principle again, taking the proposition (ii) as the value of 'A', and the proposition (iv) as the value of 'B'. By the Closure Principle, (iv), and (v), we can conclude:

(vi) *E* is committed to: *E* is committed to: *E* is committed to *C*.

It is clear that this pattern of argument can be repeated to obtain arbitrarily long iterations of 'E is committed to' preceding C.

Self-reference within (C) is wholly essential to obtaining these iterations. Suppose we had only the following information, which does not involve self-reference, about a variant encyclopaedia F:

(i') The encyclopaedia F says:

(D) Rome is Italy's capital, and the encyclopaedia F contains the information that Rome is Italy's capital.

Arbitrarily high iterations do not follow from this. F says that it says that Rome is Italy's capital; but it leaves it open (on the basis of the information given) as to whether it is committed to saying that it says it. In our little derivation above (iii) was crucial in arguing for the iterations. The analogue of (iii) does not hold in the case in which we have only (i'). (D) itself does not imply that F contains (D). (D) implies only that F contains the information that Rome is Italy's capital, and not that it contains the information (D) itself.

Now we can return to mental states and events with indexical self-referential contents. Do they generate arbitrarily high iterations? To apply and extend a remark of Barwise (1988) to our example of the encyclopaedia: commitment travels at the speed of logic, but awareness and knowledge travel at the speed of the mind. We can add that sometimes that speed is zero. Consider someone who has a genuinely reflexive awareness of his own pain. He is not merely aware that he is in pain. He rather meets the condition (vii):

(vii) he is aware that he is in pain, and he is aware this whole awareness exists.

Do arbitrarily high iterations of 'he is aware that' applied to 'he is in pain' follow from (vii)? They would if we had the following unrestricted principle:

If our subject is aware that an awareness exists, and this latter awareness involves his being  $\phi$  (for a suitably tight notion of involvement), then he is aware that he is  $\phi$ .

This unrestricted principle does not hold, if humans are not capable of arbitrarily complex states of awareness—as, to say it again, they cannot be if their minds are finite. But there is a clear sense in which the materials are available for a subject to rise to any particular level of iteration once he meets the reflexive condition (vii). For (vii) is stronger than any finite iterations of 'he is aware that' applied to contents that do not refer to the mental state of awareness in question.

So it is with full joint attention. The account of full joint attention by  $x$  and  $y$  to  $o$  that I am now suggesting is this:

(a)  $x$  and  $y$  are attending to  $o$ ;

(b)  $x$  and  $y$  are each aware that their attention in (a) has mutual open-ended perceptual availability; and

- (c)  $x$  and  $y$  are each aware that this whole complex state of awareness (a)–(c) exists.

For any particular level of iteration of ‘ $x$  is aware that  $y$  is aware that ...’, it could in principle be reached by  $x$  and  $y$ , if each co-operates and each extracts everything from the state of awareness that is required to reach that level. No finite mind can reach every level given that we are concerned with a non-dispositional form of awareness. But (a)–(c) remain very different from a non-self-referential state of awareness with finitely many iterations and a cut-off point.<sup>3</sup>

The elements of the finitely many iterations which do hold in a particular case of full joint attention also have a single common explanation: awareness with an indexically self-referential content. More generally, just as (vii) is a case of individual genuinely reflexive consciousness, this account (a)–(c) of full joint attention can be regarded as treating it as a simple two-person case of genuinely reflexive social consciousness.

Full joint attention, even if its participants cannot fully articulate its nature, is a relatively sophisticated state. Its reflexivity is possible only for beings who have some way of representing attention, mental states, and employing some form of indexical reference to mental states. Between the most primitive forms of mutual awareness required for simple co-ordinated joint actions and the mature phenomenon of full joint attention, there will be a series of increasingly rich types of mental representation, content, and operations upon them. The conceptual, as well as the empirical, investigation of this series is one of the many tasks for future work suggested by the present approach. It is, for example, well known that infants are perceptually sensitive to the distinction between those events which appear to be actions and those which do not. But looking at something, and attending to something, are actions. The ability to identify actions will be the first step into the territory in which the mature phenomenon of full joint attention is located.

Important questions also arise at this point about the relation of awareness with self-referential intentional contents to the possibility of self-involving situations, and to other self-referential mental states. It can be very tempting to use self-involving situations in describing the reflexivity of full joint attention. Such an approach raises many metaphysical issues: I pursue some of them in the Appendix to this paper, in order not to lose the main thread of the argument in the philosophy of mind proper. That thread now leads us to consider the relation

<sup>3</sup> States with indexical contents have been used by Gilbert Harman in the characterization of mutual knowledge. He observes: ‘A group of people have mutual knowledge of  $p$  if each knows  $p$  AND WE KNOW THIS, where “THIS” refers to the whole fact known’ (Harman, 1977, p. 422). Harman also makes use of self-referential intentions to analogous iterative effect in his review of Schiffer, 1988, (Harman, 1974). For more general remarks on self-referential attitudes, see Harman, 1986, esp. ch. 8. The views I defend in the text above do not imply that all intentions, theoretical conclusions, or perceptual experiences have self-referential contents.

between full joint attention as characterized here and the kind of openness involved in mutual knowledge.

## 2. JOINT ATTENTION AND MUTUAL KNOWLEDGE

I start this section by arguing that the classical account of common or mutual knowledge developed by Lewis and Schiffer does not capture the openness of some of the situations to which it has been applied in the literature.

Lewis's and Schiffer's accounts, which are in their essentials identical, are justly famous and beautiful accounts of a phenomenon with which previous writers had grappled and failed. I do not question that the Lewis–Schiffer account gives a good philosophical explanation of some cases of mutual knowledge as Schiffer defines it. I think it gives the correct explanation of mutual or common knowledge of such truths as that George W. Bush is currently President of the USA, and that people here in New York drive on the right, speak English, and use dollars and cents as money. But I think there is a class of cases to which the Lewis–Schiffer account does not apply, even though these cases nevertheless display a distinctive kind of mutual openness. This class includes even some examples by reference to which the notion of mutual knowledge was introduced in this literature. The class of cases in question is that in which there is a distinctive kind of knowledge made available by full joint attention.

To substantiate these claims, we need to look in more detail at the classical account. I will use Schiffer's formulation (points corresponding to those I will make could be set out *pari passu* for the Lewis formulation). First, here is Schiffer's development of an example, a case in which we 'Suppose that you and I are dining together and that we are seated across from one another and that on the table between us is a rather conspicuous candle' (1988, p. 31). He continues (I will change his notation for uniformity with the preceding, but otherwise this is verbatim):

Clearly I know that there is a candle on the table. So

$Kxp$ .

I also know that you know that there is a candle on the table. How do I know this? First, I know that if a 'normal' person (i.e. a person with normal sense faculties, intelligence, and experience) 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. Secondly, I know that you are a 'normal' person and I see that your open-eyed head is facing the candle... So

$KyKxp$ .

Further, I do not presume to be the only person aware of the above-mentioned law about normal people in certain circumstances; I also know that you know that normal people see things that are in their line of vision when their eyes are open, etc. And I have

seen that you see that my open-eyed head is facing the candle. So I know that you know that I know that there is a candle on the table; i.e.

$KxKyKxp$ .

Schiffer's account relies on what we can call *generating properties* (my terminology) and iterated inferences from these generating properties and their characteristics. In the example of mutual knowledge\* about the candle, the person  $x$  has the following generating property: that of being a visibly 'normal', open-eyed, conscious person who is identical with  $x$  and who, at a close distance, is directly facing the candle, and  $y$ , who has the same properties *vis-à-vis*  $x$  (cf.p. 35). The idea is that in a case of mutual knowledge between  $x$  and  $y$  that  $p$ , if  $F$  is the generating property for  $x$  (with respect to  $p$ ), and  $G$  is the generating property for  $y$  (with respect to  $p$ ),  $F$  will have two crucial characteristics:

First, being  $F$  is sufficient for knowing that  $p$ , for knowing that  $x$  is  $F$ , and for knowing that  $y$  is  $G$ . The same holds for  $G$  correspondingly.

Second, for any proposition  $q$ , if being  $F$  and being  $G$  are each sufficient for knowing that  $q$ , then both being  $F$  and being  $G$  are sufficient for knowing that sufficiency condition.

Schiffer provides a finite basis for cases of mutual knowledge\* in terms of such generating properties. His theory is then that  $x$  and  $y$  mutually know that  $p$  iff there are generating properties of  $x$  and  $y$  with respect to the proposition  $p$ .<sup>4</sup>

The hierarchy of iterations of knowledge on the part of each mutual knower is attained by the knower by inference from his knowledge that he and the other have the relevant generating properties. These iterations really do follow from this simple theory. The elegance of the approach is undeniable.

I will be arguing for three points. The first point, a negative claim, is a Thesis of Non-Necessity, to the effect that the Schifferian conditions are not required for the openness distinctive of many of the examples discussed. The second point, more positive, is that an Alternative Account can be given of a kind of openness and the knowledge it generates, an account which does not require mutual knowledge\*. The third claim is that this openness and knowledge of the sort characterized in the Alternative Account are in fact what characterize a significant range of the phenomena for which mutual knowledge was invoked by earlier writers.

The Thesis of Non-Necessity implies, then, that in some basic cases of full joint attention, there is no Schifferian generating property and there are no Schifferian iterated inferences. That is, the openness of the situation is not an inferential matter. It is rather a matter of perceptual awareness. The openness of the situation is not captured by any finite basis for inference, because it is not a (personal-level) inferential matter at all.

<sup>4</sup> This summarizes his formulation on pp. 34–5.

As a further elaboration of this Thesis of Non-Necessity, I also suggest that the cases in which there is no such inference are in a certain sense basic. The existence of these cases makes possible examples of mutual knowledge\* as characterized by Schiffer's theory.

Two people, one or both of whom may be 6 years old, can jointly attend to a candle without so much as having the conception of a normal person, let alone beliefs or knowledge about the psychological capacities of normal people. Each person may simply see that the other sees the candle.<sup>5</sup> They can also have the more complex forms of mutual awareness involved in our characterization of joint attention without engaging in inference at all, and *a fortiori* without employing a finite basis for inference.

Could it be replied that the finite basis, the relevant generating properties, are actually merely tacitly known, and that there is tacit inference from this tacitly known finite basis? Maybe so: but tacit knowledge and tacit inference, or computation, therefrom is entirely compatible with the finally attained state being a perceptual state, rather than being a personal-level inferential state. This is precisely our conception of tacit knowledge of the rules of a grammar for English. Unconscious operations which draw on the information stated in grammatical principles result in perceptual states, of perceiving a sentence as grammatical or otherwise, or of hearing it as having a certain semantic and syntactic structure. But this is not personal-level inference, under rational control. Far from being an alternative to a perceptual account of the openness involved in joint attention, merely tacit knowledge is rather one particular kind of account of how a perceptual phenomenon might arise.

How, as things actually are, do people ever come to know that in general, normal people see things that are in front of them, that normal people know this, and so forth? It seems to me that this knowledge is attained by generalization from experience with particular situations in which one sees that someone is seeing something, and in which this is wholly open in the way in which I have tried to characterize joint attention. Experience with, and knowledge of, particular situations is rationally prior to knowledge of generalizations about normal persons. It is not clear that there is a plausible alternative way, as things actually are, in which knowledge about normal people could be acquired otherwise. Further, for the relevant properties of normal people to have the distinctive characteristics of Schiffer's generating properties, one needs experience not merely of others seeing things, but of the openness of such situations. Otherwise one would not be in a position to know that normal people can attain knowledge about what normal people are like (as opposed to merely having that knowledge oneself). Common knowledge of such matters of driving conventions, and geographic and political

<sup>5</sup> Schiffer himself slips into the very natural, and in my view true, description of one person as seeing that the other sees something to be the case (1988, p. 31).

facts, all rely on common knowledge of facts about perception; which in turn relies on the phenomenon of full joint attention.

In fairness to Schiffer, I should add that his intention may simply have been to show that a finite basis for mutual knowledge\* is a genuine possibility. His concern may not have been the correct treatment of the particular example of the candle, but simply to show how mutual knowledge is so much as possible for finite minds. He did make clear one way in which mutual knowledge\* can be attained, and in that sense he gave an existence proof. All the same, we still need a correct treatment of the example of the candle, and of the openness present in various basic joint interactions with the world.

We cannot simply leave the matter with this Non-Necessity Thesis, resting content to note that not all cases of joint attention meet the conditions for mutual knowledge as characterized in Schiffer's theory. For it is highly intuitive to say that there is some kind of openness to the knowledge that there is a candle on the table when there is joint attention to the candle. That is, there is a kind of openness to that knowledge that there is a candle on the table, an openness which is captured neither by the classical theory nor by the definition of mutual knowledge\*.

At this point, in attempting to characterize this openness, we have at least two options. One option is to seek to modify the classical theory of mutual knowledge. We might try to change some of the parameters of the classical theory, in a way that respects the perceptual character of this openness. It is not easy to see how to do this. We cannot simply replace 'knows' by 'perceives' throughout Schiffer's theory. Recall that in a case of mutual knowledge that  $p$ , the generating property  $F$  for a person  $x$  had to be such that if it is sufficient for knowing an arbitrary proposition, it is also sufficient for knowing that it is so sufficient. We cannot simply replace 'know' by 'perceive' here, and obtain something true. When we consider the sort of generating properties with which Schiffer was concerned—being a 'normal' perceiver, awake, where there is light, etc.—it is implausible that when those conditions are sufficient for the subject knowing some given proposition, it is also literally perceived, rather than known, that they are so sufficient. If one says merely that this sufficiency is known, nothing follows about the availability of higher-order perceptions. Perceptual content is, as I have commented before, not closed under a priori operations on its contents. Actually this point would apply even if it were maintained that in some cases the sufficiency of the generating property for perceiving certain propositions to hold could itself be perceived. From the facts that a given thinker perceives  $A$  to be the case, and perceives  $B$  to be the case, and the fact that  $A$  and  $B$  together trivially entail  $C$ , it does not follow that our subject perceives  $C$  to be the case. I will not pursue the option of modifying the Schifferian account in any more detail, because I think that account is suited only to the inferential case. The essentially inferential character of Schiffer's account is not well-suited to phenomena of perceptual awareness.

Instead of trying to modify the classical account, I will pursue the other theoretical option. That is the option which holds that there is an openness of knowledge in cases of joint attention that is not captured by the classical theory of mutual knowledge and its inferential mechanisms. At this point I aim to develop an Alternative Account. Let us speak of *open knowledge* in its own right. In particular, consider open perceptual knowledge. We can say that

$x$  and  $y$  have open perceptual knowledge that  $p$  iff

(a')  $x$  and  $y$  both perceive that  $p$ ;

(b')  $x$  and  $y$  are both aware that their perceptions that  $p$  are mutually open-ended; and

(c')  $x$  and  $y$  are aware that they are both aware of this very awareness (a')–(c').

Two people jointly attending to a candle between them may have open knowledge that the candle is flickering, even if they do not meet Schiffer's finite basis for mutual knowledge that the candle is flickering.

The more general case of open knowledge, be it perceptual or non-perceptual, could be characterized in terms of open perceptual knowledge and what is mutually known about each other's memory and inferential procedures. This more general case would be a mixture of the perceptual character of joint attention and the inferential character of Schifferian mutual knowledge.

What are the properties of open perceptual knowledge? Open perceptual knowledge that  $p$  does not imply arbitrarily high iterations of knowledge, since, as we saw, arbitrarily high iterations of awareness are not implied by our description of joint awareness. Failure to imply arbitrarily high iterations of knowledge is not, however, the most fundamental difference between open knowledge and common knowledge. Indeed, Lewis notes that his particular formulation of common knowledge does not imply arbitrary iterations of actual knowledge or expectation, since the rationality assumptions required on his account for such a derivation become more demanding and implausible as we rise through higher levels (1969, pp. 55–6). What is distinctive of cases of open knowledge is not the absence of arbitrary iterations of knowledge, but rather the means by which such finite iterations as do hold are reached. Suppose that in the case of joint attention to the candle,  $x$  does reach this state:

$x$  sees that  $y$  sees that the candle is flickering.

Seeing that  $p$  is a form of knowledge that  $p$ : 'seeing that' is a factive mental state operator with the properties noted by Williamson (2000, pp. 34–41). It follows that

$x$  knows that  $y$  sees that the candle is flickering.

This is something that follows from the nature of  $x$ 's state: it is not a matter of  $x$  making any transition in thought. By contrast, the next and final step is a



matter of  $x$  himself making some transition in thought. Suppose that  $x$  has some appreciation that seeing is a form of knowledge. Then from that fact, some minimal inferential competence, and the last displayed sentence, we have that

$x$  knows that  $y$  knows that the candle is flickering.

There are two notable features of this derivation of second-level iterated knowledge. One feature is that the iteration has not been derived from a Schifferian finite basis. It has not been reached by any inferences about 'normal' people. It has instead been extracted from a complex state of awareness present in a case of joint attention. The openness of a situation of joint attention consists in the facts, perceptual facts, given in our characterization of joint attention. Open knowledge is a by-product of that perceptual openness.

The other notable feature of the derivation is that it requires  $x$  to appreciate that seeing something provides a way of gaining knowledge about that object. Autistic children, even able ones, do not in general have this appreciation. A study by Josef Perner, Uta Frith, Alan Leslie, and Susan Leekam (1989) suggests that only one third of able autistic children realize that seeing something provides a means of coming to gain knowledge about that thing.<sup>6</sup> If their figure is roughly right, it follows from their work and the present account that most autistic children will not be capable of attaining open knowledge. Even if an autistic child sees that someone else is looking at something, he does not thereby gain knowledge of what the other knows. This state of affairs may greatly reduce the interest, for the autistic child, of situations in which he sees that he and another are seeing the same thing. Actually, we can distinguish two points at which there may be failure to make a transition to the attribution of knowledge to the other person. Perner *et al.*'s questions to autistic children were framed in terms of what the other person was looking at (1989, p. 693). To reach open knowledge by the route I have been describing, one must first be able to make the move from information about which objects the other is seeing or looking at to information about what the other sees to be the case; and then one must be able to move from this information to an attribution of knowledge. Failure could occur at either of these two points.

### 3. WIDER APPLICATIONS

My third claim was that there is a range of phenomena for which the Alternative Account, in terms of open knowledge and its source in joint attention, does better than mutual knowledge. Before generalizing, I take first the openness of linguistic communication. Suppose I say to you 'It's time for lunch.' My utterance is

<sup>6</sup> See esp. Perner *et al.*'s (1989) discussion of knowledge-formation tasks.

successful if we are jointly aware that I am saying that it is time for lunch. More generally, the paradigm of a successful indicative utterance in which one says that *p* is an utterance of which utterer and audience are jointly aware that it is a saying that *p*. A similar point applies *pari passu* for moods other than the indicative. The utterer aims to bring about a state of joint awareness whose content involves the meaning or, better, the intentional content of the utterance itself.

It is essential for this account that the intentional content of the utterance be part of the representational content of the perceptual experiences (in this case auditory experiences) of utterer and listener. If the assignment of meaning to the utterance were merely a matter of personal-level inference on the part of the hearer, then joint awareness that I've said that it's time for lunch would not be perceptual. But it seems that in the paradigm cases of successful communication, it is so.

The openness of communication does not require the inferential structures in Schiffer's account of mutual knowledge. The openness of my communication consists rather in the fact that you and I have full joint awareness that I am saying that it's time for lunch. We equally have open knowledge that I said it's time for lunch.

I don't need to have any beliefs about what normal people in our linguistic community do, or how they interpret utterances, to perceive you as saying that it's time for lunch. Nor do I need such beliefs for us to be jointly aware of this fact about what you are saying. In fact, in the cases in which I do have such knowledge, it seems to me once again to be based on experience of communications with relatives and acquaintances which already involve the kind of joint awareness I am identifying, in advance of any knowledge about the community. Even if some of my beliefs about what the wider community would mean by utterances of certain expressions are false, that need not prevent me from hearing your utterance correctly, and from our having full joint awareness of my saying that it's time for lunch. Just as false beliefs about the mechanisms of perception do not prevent one from having ordinary perceptual knowledge about the world—since that knowledge does not rest on beliefs about the nature of perceptual mechanisms—so also having false beliefs about language in the community does not prevent one from being aware of what someone has said on a particular occasion.

Sometimes, when an expression or a surface syntactic structure has to be disambiguated, beliefs about one's circumstances, and the likely topic of conversation, will affect how one hears one's interlocutor's utterance. But this does not mean that knowledge of meaning is purely inferential. The point is analogous to perceptual cases outside those of linguistic communication. Your knowledge of your peaceful circumstances means that you will discount an apparent perception as of a machine-gun firing as incorrect. You may realize that, say, some ball-bearings have tumbled in rapid succession on a metal surface; and you may (but need not) come to hear the continuing event as such a tumbling. Perception remains distinct

from judgement and belief; but it is false to say that it cannot be influenced by them, or by hypotheses entertained by the thinker. Sometimes also, disambiguation itself is not a matter of conscious inference. The continuing context disambiguates without any entertaining or thought of alternative readings. The end state of the understander is still something perceptual, rather than merely judgement reached by inference.

I have concentrated on particular single instances of full joint attention; but in fact full joint attention is rarely a matter of one-off single events. Joint attention and contact attention between two people are commonly extended over time. Examples are ubiquitous, and so important, that one is inclined to say that their possibility is part of what it is to be distinctively human. Cases range from temporally extended games and interaction with a child, to adult interactions, both conventional and non-conventional.

In an extended encounter between two people involving joint attention and contact attention, we can speak of the mutual world which is created between the participants. This mutual world involves the events which are jointly attended to, the participants' relations to them, the development of both of these over time, and what is known at each stage about what has happened earlier in the mutual world of joint attention. Nagel's well-known paper 'Sexual perversion' (1969) was right to emphasize the role of iterated psychological states in mutually open sexual interactions between two people. Here too I believe the phenomena are best characterized in terms of full joint attention and full contact attention, rather than in terms of common knowledge. In such interactions too, a joint-attentional world is created between the participants.

Extended discourse between two people also creates such a mutual joint-attentional world. Some of the matters that David Lewis calls score-keeping in a language-game concern the created world of joint attention constructed and developed in the discourse to which the participants are jointly attending. Language has multiple special features; but I suggest that the openness of linguistic communication should be seen as a special case of the philosophically prior and more general phenomenon of the openness of joint attention.<sup>7</sup>

<sup>7</sup> In discussing the openness of linguistic communication, Simon Blackburn writes: 'The ideal of full openness is more simply captured if we just add the want that nothing about my wants be concealed' (1994, p.115, in a section entitled 'Openness and communication'). Blackburn's discussion is in several respects consonant with the claims of the present paper; but I would add that the best way for such a want as he cites to be fulfilled is to create a situation of full joint attention. When an intention to produce such a situation is fulfilled, a certain openness will be present as intended, even if the agent does not have self-referential wants of the kind Blackburn discusses. Blackburn also writes, congenially, 'But in the restaurant is it really plausible to say that when our eyes meet in full mutual awareness, I (and she) have an endless stock of wants?' He says that his condition that the subject wants no concealment of his wants describes the situation better. I would say that this still does not pick out a distinctive state of awareness, whatever the beliefs and desires of the participants. What distinguishes our two people in the restaurant is that they have open perceptual awareness that they are attracted to each other.

## APPENDIX: REFLEXIVITY, SELF-INVOLVING SITUATIONS AND SELF-REFERENTIAL STATES

This Appendix discusses the relation between the above treatment of full attention and two other approaches found in the extensive but scattered literature on other issues about mutual psychological properties. The first of these other approaches makes use of self-involving situations; and it plunges us into metaphysical issues.

*Self-Involving Situations*

It can be tempting to elucidate the openness of full joint attention in terms of self-involving situations thus: when  $x$  and  $y$  are engaged in full joint attention to  $o$ , there is a situation  $S$  such that:

- (si)  $S$  is a situation in which  $x$  and  $y$  are attending to  $o$ ;
- (sii)  $S$  is a situation in which  $x$ 's and  $y$ 's awareness that they are both attending to  $o$  is open-ended; and
- (siii)  $S$  is a situation in which  $x$  and  $y$  are perceptually aware of  $S$ .

Such a situation  $S$  is not merely self-involving. It is more specifically what we can call *psychologically self-involving*, in the sense that it involves a psychological relation to itself. Self-involving situations have been invoked by several writers, in a range of disciplines from psychology, philosophy, and linguistics, to account either for mutual knowledge, or for weaker versions of mutual mental states.<sup>8</sup> One of the most vigorous, and formally creative, proponents of this style of approach, Jon Barwise, went so far as to write that 'Shared understanding in all its various guises (mutual belief, common knowledge, public information) rests on circular, or at least non-wellfounded situations. In as much as there are assumptions of such shared understanding throughout game theory, law, communication theory, and the like, we are constantly caught in non-wellfounded situations' (1989, p. 198).

Psychologically self-involving situations also have the power to give rise to iterations of attitudes. Once again, for the sake of illustration, we start with a state other than awareness, to make the formal structures clear. Suppose we have a person—John—and a situation  $S$ , which consists in the holding of (sa) and (sb):

- (sa)  $p$
- (sb) John knows that  $S$  exists.

Let us write  $S \models A$  for 'A holds in situation  $S$ ' (as in 'A is a fact of situation  $S$ ' in the way this is used in the situation theory of Barwise and Perry (1983; see also Barwise, 1989)). Suppose also that we have the following Principle K holding of John's knowledge:

(Principle K) If John knows that  $S$  exists and  $S \models A$ , then John knows that  $A$ .

From (Principle K), (sa), and (sb), it follows that

- (1) John knows that  $p$ .

Hence

- (2)  $S \models$  John knows that  $p$ .

<sup>8</sup> There is significant and illuminating use of self-involving situations in Clark and Marshall, 1981; Barwise, 1988; Sperber and Wilson, 1995, p. 42.

So from (Principle K), (sb), and (2), we have

- (3) John knows that John knows that  $p$ .

Hence

- (4)  $S \models$  John knows that John knows that  $p$ .

And so forth, for arbitrarily many iterations of 'John knows that' applied to  $p$ . All of these follow from Principle K and the individuation of the psychologically self-involving situation  $S$ . The individuation of  $S$  together with Principle K are jointly clearly stronger than any finite list of iterations of 'John knows that' applied to  $p$ . No finite set of such iterations, however long, has the same implications as the self-involving situation  $S$  when combined with Principle K. So arbitrary iterations can be obtained from a psychologically self-involving situation, in the presence of a suitable principle governing the psychological relation in question.

Now we can return to full joint attention and its situation-theoretic characterization in conditions (si)–(siii). Actual perceivers are not like ideal knowers. Consider the principle

- If  $x$  perceives situation  $S$ , and  $S \models A$ , then  $x$  perceives that  $A$ .

We cannot expect such a principle to hold without restriction. When  $S$  is psychologically self-involving, as in our characterization of full joint attention, this principle would imply the existence of arbitrarily high iterations of 'perceives that' in  $x$ 's perceptual states. Since these states—unlike beliefs conceived as inexplicit in Schiffer's account—are not dispositional, this principle would imply the existence of actual occurrent perceptual states that full joint attenders do not in fact enjoy. Ordinary perceivers go only so far in extracting what it is in principle possible to extract from a perceptually self-involving situation. So the clauses (si)–(siii) do not commit us to saying that  $x$  and  $y$  actually have arbitrarily complex perceptual states with contents like 'He perceives that I perceive that he perceives...'. When  $x$  is perceptually aware of the state  $S$  described in clauses (si)–(siii),  $x$ 's perceptual system has the information to compute correctly from states it is already in to new contents, such as 'y is aware of a state in which I'm aware of a state in which he's attending to o'. But it does not follow from the fact that  $x$  has the information to compute, subpersonally, this content, that he actually carries out the computation. On all of these most recent points, the advocate of self-involving situations for characterizing full joint attention can say much the same as I said in the main text, in defending the treatment in terms of mental states and events that are self-referential. So why not describe all the phenomena with which this paper is concerned by using the apparatus of self-involving situations?

The main reason must be concern that self-involving situations are not metaphysically legitimate. They seem to involve an unacceptable regress of individuation. Anyone who has been brought up on the iterative hierarchy as the intended model of ZF set theory will have been trained to find self-involving situations suspect. No doubt training can induce acceptance of almost anything—but if we want a rational articulation of the concern, it can be formulated as the principle that the world itself must be well-founded. I call this position 'insistence on objective well-foundedness'. Opponents of objective well foundedness have developed illuminating formal theories of what self-involving situations would have to be like, and of what is involved in commitment to them. Peter Aczel's (1988) theory of non-well-founded sets is a particularly elegant exposition of the foundations of such a treatment, and of its commitments. Aczel shows that the existence of non-well-founded sets is implied

by nothing more than the axiom that every graph has a unique decoration. Valuable as this formal development is, I do not think that by itself it can answer the doubts of someone who insists on objective well-foundedness. The simplest case of a one-element graph which generates a set which is a member of itself seems to involve something which is individuated in terms of itself (Aczel, 1988; Barwise and Moss, 1996, ch. 10). This will concentrate on, rather than answer, the doubts of those who insist on objective well-foundedness.

I will be offering some arguments and theses in favour of the conjecture that reality is well-founded. One of the major tasks facing anyone who holds this position is to explain how, consistently with that position, there can exist all the examples which Barwise cites and which he says require the existence of self-involving situations (1989, pp. 194–8; Barwise and Moss, 1996, pt. II, esp. ch. 4).

The first step in elaborating this position is to draw a distinction between *eliminable* and *ineliminable* self-involving situations. An eliminable self-involving situation is one whose existence can be fully explained in terms of situations or other entities that are not self-involving. The person who insists that reality must be objectively well-founded should not have any quarrel with the existence of eliminable self-involving situations. How could a self-involving situation be eliminable? Here is one way, which draws upon the materials of the preceding discussion. Some self-involving situations consist in the occurrence of self-referential mental states or events. In such cases, the self-involving situations can be obtained from self-referential mental states by what I call ‘an inside-out transformation’. Suppose a person  $x$  has the following properties:

$x$  is aware that  $p$ ; and

$x$  is aware that this whole awareness exists.

I suggest that we can, quite generally, transform such a state of affairs into the following true statement about a self-involving situation:

There is a situation  $S$  which is constituted by the facts that

$x$  is aware that  $p$ , and

$x$  is aware that  $S$  exists.

A self-involving situation which is attainable by such a transformation is in the nature of the case eliminable in the above sense. Its existence is wholly consistent with reality being objectively well-founded. Provided a self-involving situation is attainable by an inside-out transformation, we can legitimately use it in characterizing full joint attention and other phenomena of mutuality, whilst continuing to insist that reality is well-founded. The main text of the present paper can then be regarded as doing the following. If the self-involving states usable in characterizing full joint attention are obtainable by an inside-out transformation, it ought to be possible to specify the well-founded states from which they are so obtainable. The main text of the paper then aims to answer the question: which are those well-founded states? Certain mental states with indexically self-referential contents provided the answer.

If this is correct, then the permissibility of using self-involving situations in the description of these psychological phenomena does not by itself legitimize self-involving situations that are not reachable by use of inside-out transformations. These psychological phenomena cannot be used in support of the claim that the world is not well-founded. It would take some real phenomenon describable only by self-involving situations that are ineliminable to refute objective well-foundedness.

These considerations can be generalized to provide a case in support of objective well-foundedness. The case can be presented in a series of four theses, of which these are the first three:

Thesis 1: There is a range of types of well-founded states of affairs that generate legitimate, eliminable self-involving states of affairs.

Thesis 2: Indexically self-referential mental states form just one of these types.

Thesis 3: What all these types have in common is that from well-founded states of affairs, they generate the infinitary characteristics distinctive of self-involving situations.

One example of a well-founded state of affairs which generates eliminable self-involving states of affairs is Schiffer's classical inferential account of mutual knowledge. He gave a finitary, inferential explanation which applies to the case of mutual knowledge between you and me that people in the USA drive on the right, in terms of the generating properties we noted in Section 2 above. The conditions cited in Schiffer's inferential explanation are wholly well-founded. They are also the conditions in virtue of which there exists a self-involving situation  $S$  with the following properties:

$S \models$  People in the USA drive on the right.

$S \models$  I know  $S$ .

$S \models$  You know  $S$ .

(Here I follow the style of characterizing mutual knowledge found in Barwise, 1989, and in Barwise and Moss, 1996, p. 49.) The relevant infinitary consequences of this self-involving situation are also consequences of Schiffer's finitary characterization. There is, then, a legitimate self-involving situation that can be mentioned in the explanation of some cases of mutual knowledge; but it is generated by conditions—those formulated by Schiffer—which are wholly well-founded.

A non-psychological case cited by Barwise as non-well-founded is an example due to John Perry. I quote:

Imagine two parallel mirrors of the same size, facing each other, one  $A$  with an 'X' painted on it, the other  $B$  with an 'O'. This is a simple finite physical situation which is, in some sense circular. We can think of it as three situations, a situation  $s$  with two sub-situations  $s_A$  and  $s_B$  the scenes reflected in  $A$  and  $B$ , respectively. In  $s_A$  we have the facts that  $B$  has an 'X' and that  $B$  reflects  $s_B$ , while in  $s_B$  we have the symmetric facts about  $A$ . The facts of  $s$  are those of  $s_A \cup s_B$  plus the facts that  $A$  and  $B$  are parallel and facing one another. (Barwise, 1989, p. 194).

This description is highly intuitive: but the self-involving situations used in the description are also eliminable. They exist in virtue of the wholly well-founded spatial relations between  $A$  and  $B$ , the markings on them, and the laws of reflection. These well-founded conditions have infinitary consequences about reflection (if we prescind from the minimum size of particle which can reflect light, and other minima). Nothing here involves ineliminably self-involving states of affairs.

More generally, I suggest that what holds of the mirror example holds also of the other examples cited in support of the claim that the world is not well-founded. I suggest:

Thesis 4: Each of the examples of 'circular' situations cited by Barwise (1989) and by Barwise and Moss (1996) is either

(a) not really self-involving, or

(b) it is a self-involving situation generated by one of the types of well-founded situation mentioned in Thesis 1, and so is a case of merely eliminable self-involvement.

I won't enumerate every single example developed in the works by Barwise and by Barwise and Moss. I will rather take a few as exemplars, and aim to give responses available to one who holds that reality is well-founded, responses which can be transferred to other cases.

Some of the cases which are described as cases of circularity do indeed meet that description, but do not amount to the postulation of self-involving situations. Barwise and Moss describe, for instance, one way of solving a particular probability problem, that of finding the probability that the first time we get heads in a series of tosses of a fair coin, it will be on an even-numbered flip. An elegant solution to this problem involves characterizing the unknown,  $x$ , as identical to a number picked out by a condition involving  $x$ . (Barwise and Moss, 1996, p. 53).<sup>9</sup> This gives an equation with  $x$  on both sides, which can be solved.

There are no self-involving situations in this case at all. For there to be a complex condition on  $x$  that involves  $x$  itself, and which is uniquely satisfied, is not for any entity or state of affairs to be individuated in terms of itself. It is an important point that a specification of an object  $x$  may involve  $x$  itself, and may uniquely fix  $x$ . This is something that is relied upon in functional specifications of states in the philosophy of mind, and in the individuation of concepts in some theories of concepts (Lewis, 1970, 1972; Peacocke, 1992, esp. ch. 1). But it does not require self-involving situations or non-well-founded states of affairs.

The same seems to me to be true of self-reference, and of reference by tokens or events to larger tokens or events of which they are a part. Some words in a sentence may refer to the whole sentence of which they are a part, as in Barwise's example of an airline announcement which concludes, 'This announcement will not be repeated'. Barwise and Moss say that 'Such examples show that the relation "refers to" is circular, that is, that things sometimes refer to themselves, or to other things that refer back to them' (1996, p. 56). This falls short of showing that the world is not well-founded. The token utterance-events in the airline announcement—the utterance of 'this announcement' and the utterance of the whole announcement—are neither of them individuated in terms of themselves, nor circularly in terms of each other. No utterance-token, or inscription, or even an event of reference in thought, is individuated in terms of itself, in the way that the non-well-founded sets of Aczel's theory are individuated in terms of themselves.

But perhaps we are looking in the wrong place for the non-well-founded. Barwise and Moss's remark is explicitly about the relation of reference: so is that relation itself not well-founded? For a relation not to be well-founded is for it to be individuated in terms of itself. To establish that a relation has that property, one has to show something about what it is in general for the relation to hold, and in particular to show that the conditions for its holding can be given only in terms of that very same relation itself. Barwise and Moss do not do that: it is entirely consistent for a well-founded relation to hold between an entity and itself, or an entity and some other entity of which it is a part. Further, to be an objection to the well-foundedness of the world, it would have to be shown that any self-involving

<sup>9</sup> The example is from Donald Newman.



characterization of the reference relation is ineliminable. The arguments in these works of Barwise and of Barwise and Moss have not done that either.

Several of Barwise's (1989) examples concern mental events and states. He discusses Descartes's thought 'I think therefore I am', and writes of it that 'it is a mental act that comprehends itself as a constituent' (p. 194). Barwise's point is that it is Descartes's thinking of this very thought that convinces Descartes, in the face of the doubt, that he exists. But this is then a case of the sort we discussed in the first section: it is a mental event with an intentional content one of whose components refers to the mental event itself. In fully explicit form, the content is 'I think in this very thinking, therefore I am' (indeed, this is very close to the later exposition in Barwise and Moss, 1996, p. 51). But, once again, for a mental event to have an intentional content one component of which refers to that mental event is not for anything in the world to be individuated in terms of itself. In my judgement, the same applies to Barwise's other psychological examples.<sup>10</sup>

Finally, a case which is more challenging and demands a more detailed response is Barwise's (1989) example of the situations involved in the game of poker. Barwise notes that certain parts of a play of a game of poker are public information, and that what is public—such as the information that 3♣ is an 'up' card—are circular (p. 197). What is public is a situation in which not merely is it seen by all players that 3♣ is an 'up' card, but they all see that it is, they all see that they all see that it is, and so forth. An advocate of self-involving characterizations would say that there is a self-involving situation *S* which consists in

- (i) each player seeing 3♣ face up on the table, and
- (ii) each player seeing this very situation *S*.

This self-involving situation will then have the familiar iterative consequences which we require if we are to describe the public part of this play of the game fully and accurately.

As before, I do not object to this description of the situation, but I do hold that this self-involving situation is eliminable. The well-founded state of affairs in virtue of which it holds is this. Each player *x* of the game is in the following state:

- x* sees that 3♣ is face up on the table, and
- x* sees that this whole seeing is seen by the other players.

This well-founded state has the relevant iterative consequences. To give the barest outlines of how the consequences follow, suppose we have as a premiss that this seeing by *x* is a seeing that *p*. We know that this seeing is seen by the other players; so (under the circumstances of the present case), the other players see that in this seeing *x* sees that *p*. We can apply this form of reasoning to conclude that the other players see that in this seeing *x* sees that 3♣ is face up on the table. But if the player *x* is computationally competent, he sees that these premisses hold and can compute soundly from them, and will thereby come to see, in this seeing, that the other players see that he sees that 3♣ is face up on the table; and so forth. As before, the indexically self-referential seeing has more iterative consequences than any finite series of iterations without self-reference. Barwise was right to emphasize this feature of the situation; but we can have that feature, and explain its presence, without postulating ineliminably self-involving situations.

<sup>10</sup> Barwise cites what he calls a Gricean intention that that very intention be recognized. This too is a case of a mental state with an intentional content one of whose components refers to the mental state.

*Fixed-Point Characterizations*

As we noted, a principle which characterizes a notion informatively in terms of itself is what logicians call a fixed-point principle. One approach developed by some theorists of mutual notions is to try to use fixed-point principles to individuate those very notions themselves. In the case of mutual knowledge, for instance, we can say that  $x$  and  $y$  mutually know that  $p$  iff

- $x$  and  $y$  know that  $p$ , and
- $x$  and  $y$  know that it is mutual knowledge that  $p$ .

In his paper 'Three Views', Barwise noted that this fixed-point characterization will not be equivalent to the characterization in terms of repeated finite iterations of knowledge when  $x$  and  $y$  have limited logical abilities (1988, p. 378). Joint attention under the approach I have been suggesting also supports a fixed-point principle.  $x$  and  $y$  are jointly attending to  $o$  iff:

- $x$  and  $y$  are attending to  $o$ ;
- $x$  and  $y$  are aware that this attention is open-ended; and
- $x$  and  $y$  are each aware that they are jointly attending to  $o$ .

The above discussion shows that we do not require ineliminable self-involving situations to explain how it can be that a particular fixed-point principle holds. Fixed-point characterizations, taken in themselves, are of course not well-founded; but in all the cases discussed here, the fixed-point characterizations of mutual psychological states that are true hold in virtue of psychological states and events that are well-founded.

## REFERENCES

- ACZEL, PETER (1988), *Non-Well-Founded Sets*, CSLI Lecture Notes 14. Stanford, Calif.: CSLI.
- BARWISE, J. (1988), 'Three views of common knowledge', in M. Vardi (ed.), *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*. Los Altos, Calif.: Morgan Kaufmann, 365–79.
- (1989), *The Situation in Logic*, CSLI Lecture Notes 17. Stanford, Calif.: CSLI.
- and MOSS, L. (1996), *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena*, CSLI Lecture Notes 60. Stanford, Calif.: CSLI.
- and PERRY (1983), *Situations and Attitudes*. Cambridge, Mass.: MIT Press.
- BLACKBURN, SIMON (1994), *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Oxford University Press.
- CAMPBELL, JOHN (2002), *Reference and Consciousness*. Oxford: Oxford University Press.
- CLARK, H., and MARSHALL, C. (1981), 'Definite reference and mutual knowledge', in A. Joshi, B. Webber, and I. Sag (eds.), *Elements of Discourse Understanding*. Cambridge: Cambridge University Press, 10–63.
- HARMAN, GILBERT (1974), review of Schiffer, 1988, *Journal of Philosophy*, 71: 224–9.
- (1977), review of Jonathan Bennett, *Linguistic Behaviour, Language*, 53: 417–24.
- (1986), *Change in View*. Cambridge, Mass.: MIT Press.
- LEWIS, DAVID (1969), *Convention: A Philosophical Study*. Cambridge, Mass.: Harvard University Press.

- LEWIS, DAVID (1970), 'How to define theoretical terms', *Journal of Philosophy*, 67: 427–45.
- (1972), 'Psychophysical and theoretical identifications', *Australasian Journal of Philosophy*, 50: 249–58.
- NAGEL, THOMAS (1969), 'Sexual perversion', *Journal of Philosophy* 66: 5–17.
- PEACOCKE, CHRISTOPHER (1992), *A Study of Concepts*. Cambridge, Mass.: MIT Press.
- PERNER, JOSEF, FRITH, UTA, LESLIE, ALAN, and LEEKAM, SUSAN (1989), 'Exploration of the autistic child's theory of mind: knowledge, belief and communication', *Child Development*, 60: 689–700.
- SCHIFFER, STEPHEN (1988), *Meaning*. Oxford: Oxford University Press.
- SPERBER, D., and WILSON, D. (1995), *Relevance*, 2nd edn. Oxford: Blackwell.
- WILLIAMSON, TIMOTHY (2000), *Knowledge and its Limits*. Oxford: Oxford University Press.