Joint attention, triangulation and radical interpretation: A problem and its solution

Brinck, Ingar

Published in:
Dialectica

2004

JOINT ATTENTION, TRIANGULATION AND RADICAL INTERPRETATION: A PROBLEM AND ITS SOLUTION[††]

Ingar BRINCK[†]

ABSTRACT

By describing the aim of triangulation as locating the objects of thoughts and utterances, Davidson has given triangulation the double role of accounting for both the individuation of content and the sense in which content necessarily is public. The focus of this article is on how triangulation may contribute to the individuation of content. I maintain that triangulation, interpreted in terms of joint attention, may serve to break into the intentional circle of meaning and belief, yet without forcing us to renounce the claims concerning the interdependence of meaning and belief and the irreducibility of meaning.

1. The role of triangulation for radical interpretation

Linguistic communication requires that speaker and hearer share something in order that they understand each other. That which is shared is typically thought to be linguistic conventions or some kind of rules. However, neither conventions, nor rules seem necessary for understanding. Donald Davidson replaced conventions by intentions and triangulation in his account of meaning and communication (1984). He suggested that understanding only require sharing the world in which interpretation takes place.

The aim of this article is to critically examine the nature and role of triangulation in the context of radical interpretation.[1] I will focus on triangulation as a way of finding non-semantic evidence for what words mean. In the first two sections, the concept of triangulation is introduced. The third section offers an argument in support of the view that meaning is socially constituted. In the fourth section, I present a problem for triangulation that takes the form of a dilemma. Triangulation, as Davidson described it, is either pre-cognitive or propositional (*cf*. Davidson 1997b, 128). In neither case can it in a satisfactory way contribute to determine content. In the following section, I suggest that reinterpreting triangulation in terms of joint attention will solve the problem. Joint attention is not a purely causal process, nor does it necessarily involve propositional thought. Section 6 presents an analysis of joint attention, and explains how joint attention may contribute to determine content by paving the way for language entry. In the last section, I explain how speakers during the process of joint

---

[††] I would like to thank Louise Röska-Hardy for inspiration, and Martin Jönsson and Peter Pagin for most helpful comments. Thanks also to the editorial assistant and an anonymous referee for suggestions that have significantly improved the clarity of the article. Work on this article was in part supported by the project *Language, Gestures, Pictures* at Lund University.

[†] Department of Philosophy, Lund University, Kungshuset Lundagård, SE 222 22 Lund, Sweden; ingar.brinck@fil.lu.se

[1] The article primarily concerns Davidson's later work published from the mid80ies and onwards. During this period triangulation acquired a central place in his account of interpretation.

attention may express and understand communicative intentions by having at most intentional states about states of attention. Having higher-order intentions, or intentions about intentions, is not necessary for joint attention.

Davidson held that the speaker's intention to be interpreted in a certain way provides the norm for her linguistic behaviour. In order to communicate, the speaker must intentionally make herself interpretable to a hearer. She does so by providing the clues required to arrive at the correct interpretation of her utterance (Davidson 1987, 28; Davidson 1991a; Davidson 1992). Thus, the speaker prepares for the interpreter's matching sentences of his or her own language (or idiolect) with sentences in the speaker's language (or idiolect). Furthermore, the hearer should be aware that the speaker intends the hearer to assign truth conditions to the speaker's utterances. Otherwise, interpretation will not get off the ground.

The basic connection between words and things is established by causal interaction between people and the world, and this connection constrains interpretation. Meaning is partly determined by the circumstances in which speakers learn and use linguistic expressions (Davidson 1987, 28f). Speakers, if they want to be understood, cannot mean whatever they want by the words they use, but must use language in a way that makes it possible to interpret. This is a consequence of the requirements of learnability and interpretability; that is, of making oneself understood (Davidson 1987, 28).

The hearer's assignment of truth conditions to the sentences that the speaker utters proceeds by radical interpretation (Davidson 1974). Radical interpretation occurs when the speaker is interpreting utterances by simultaneously delivering a theory of belief and a theory of meaning for which conclusive evidence cannot be had. In principle, the interpretation of a particular sentence demands the construction of a comprehensive theory for the interpretation of a potential infinity of sentences.

Radical interpretation is involved in all cases of linguistic understanding, whether it concerns a speaker of the same natural language as the interpreter, a speaker of some unknown language, or the learner of a first language (Davidson 1973; Davidson 1986b). Interpretation is indeterminate, because "alternative ways of stating the facts remain open" (Davidson 1973, 173). Davidson writes:

> Because there are many different but equally acceptable ways of interpreting an agent, we may say, if we please, that interpretation or translation is indeterminate, or that there is no fact of the matter as to what someone means by his or her words (1991a, 161).

All possible evidence could not determine a unique theory.

In radical interpretation, meanings are attributed to utterances on the basis of inferences about beliefs, and beliefs are attributed on the basis of inferences about the meaning of utterances. In order to break into the intentional circle, to separate meaning and belief, the interpreter must be charitable and make use of the principles of coherence and correspondence (Davidson 1975; Davidson 1991a).

The principle of coherence suggests that the individual speaker's thoughts and utterances are logically consistent. It also involves the idea that the speaker's actions are reasonable in the light of her beliefs and desires. The principle of correspondence pertains to the speaker's reactions to what happens in the world around her. It urges the interpreter to assume that the responses of interpreter and speaker correspond, that is, that interpreter and speaker are cognitively similar. It endows the speaker with what the interpreter takes to be true beliefs about the world. Davidson writes:

> All creatures classify objects and aspects of the world in the sense that they treat some stimuli as more alike than others. The objective criterion of such classification is similarity of response. Evolution and subsequent learning no doubt explain these patterns of behaviour (1991a, 159).

Together, the two principles assign on normative grounds rationality to the speaker and thus maximise intelligibility (Davidson 1991a, 162).[2]

But, on their own, these principles will not help the interpreter to actually perform the required matching between her own and the speaker's sentences in the context of utterance. They provide a starting-point and a guide as to how the interpreter should conceive of the speaker's overall cognitive make-up. But they will not provide the interpreter with specific clues concerning which meanings should be assigned to which utterances. Here is one point at which the notion of triangulation comes into play, in allowing for explicit language entry.[3]

Triangulation constitutes another way of breaking into the intentional circle than by using the principle of charity. Davidson (1974) suggests two strategies for breaking the circle, one of which consists in finding non-semantic evidence for the adequacy of a theory of interpretation. This strategy was described above. The second one consists in finding non-semantic evidence for what words mean. The theory of triangulation can be seen as an attempt to spell out the second strategy.[4]

Triangulation depends on knowledge concerning how perception yields the contents of belief (Davidson 1998, 89), and involves a first person (the speaker), a second person (the interpreter), and an object located in common space. By triangulating, or interpreting each other's responses to the object, the speakers attempt to converge on a similar conceptualisation of the object. Meanings are attributed on the basis of behavioural evidence concerning how the speakers categorise objects. Triangulation is necessary, albeit not sufficient, for establishing which concepts a subject has (Davidson 1992, 119). It relates the cause of an utterance to distal stimuli. It is the social character of distal stimuli that makes them the relevant determiners of content (Davidson 1997b, 128).

Repeated triangulation is a way of creating shared practises of categorisation, thus making it possible for the interpreter to judge whether the speaker's responses are correct or not. The concept of error is grounded in social interaction (Davidson 1999). Whether shared practises really exist is a moot question. According to Davidson, it does not make sense to ask for standards anchoring categorisation. Speakers can only rely on mutual interpretation (Davidson 1997a, 83).

2. How triangulation works

Two necessary conditions for linguistic communication were mentioned above. These are that the speaker make herself interpretable, and that the hearer is aware that the speaker intends for

---

[2] By this time, Davidson had given up the stronger idea of maximising agreement, as opposed to maximising intelligibility.

[3] The term "language entry" concerns language acquisition of a first natural language (such as English or French) and of any other subsequent natural language, and also language acquisition between speakers of the same natural language, since different theories of interpretation may be required for the different idiolects of speakers of the same natural language (*cf.* Davidson 1986b).

[4] Davidson also uses triangulation to argue for the public nature of meaning against empiricist theories of meaning and scepticism (1992, 262f). He claims that without an interpreter it is impossible to say what a speaker is responding to, and what the speaker's concepts are about. Because social factors play a role for the fixation of meaning, meaning can be public, i.e., accessible to all competent speakers of a language (Davidson 1987, 28).

her to interpret the speaker's utterances. An additional condition becomes pertinent in relation to triangulation. It pertains to how the speaker conceives of the external world. The speaker's responses to an event must be knowingly and intentionally responses to a specific stimulus, that is, to what Davidson calls "an aspect of a cause" (1992, 120). This means, according to Davidson, that the speaker must have the concept of the stimulus in question (*ibid*.).

For a cause to be a common cause, speaker and interpreter must conceive of it in a similar way, or, as Davidson puts it, share an aspect of the cause (1991b, 201). An aspect of a cause concerns that or those properties of the cause to which a speaker or interpreter responds or reacts (Davidson 1991b, 202; Davidson 1992, 120). The notion of a common cause is related to the assumption that language users naturally find events similar, an assumption that supposedly follows from the principle of correspondence together with facts about evolution and learning.

Triangulation depends on three similarity patterns concerning the speakers' responses to incoming sensory stimuli (Davidson 1991b, 203; Davidson 1992, 119). The speaker finds *x*s similar, the interpreter finds *x*s similar, and the interpreter finds the speaker's repeated responses in the presence of *x*s similar. The interpreter groups together the similarity responses of the speaker in order to find something of an equivalence group with her own responses. She then ascribes the speaker the concept of that object to which her own responses correspond. Consequently, the common cause is located in the intersection of the respective links of speaker and interpreter to the stimuli that cause their responses. The common cause occurs where the responses of speaker and interpreter meet. "Communication begins where causes converge", asserts Davidson (1986a, 318).

The basic relations in triangulation, that is, the speakers' responses to the external situation and to each other's responses, are causal (Davidson 1991b, 201). Had the relations been merely causal, thought contents would have to be determined purely causally. But that is not the case. The aspects of the cause that determine content are intentionally and consciously shared by speaker and interpreter.

According to Davidson, to have a concept of an object is to recognise the existence of a triangle, with oneself, a creature that is similar to oneself, and a spatiotemporally located object at the apexes (1992, 120). Thus, to communicate, the speaker must know that she is taking part in triangulation, and consequently recognise the existence of the interpreter, as a creature with intentional states, as well as of objects in a common space.

The content of an utterance is determined by an aspect of that which causes the speaker to hold the utterance true. In order for both speaker and interpreter to hold the utterance true, as a result of its being caused by the same entity, they must share the concept of truth, and conceive of the world similarly. It is the sharing of aspects by having similar true beliefs about the world, which makes possible successful linguistic communication. Common causes occur in an intersubjective world.

Davidson claims that without triangulation, words and concepts lack content. On the other hand, without concepts there could be no proper triangulation. The possibility of triangulation depends on that speaker and audience have concepts of each other and about entities in what Davidson calls a common world. Knowledge of another mind, being necessary for triangulation, is "essential to all thought and all knowledge" (Davidson 1991a, 160). The knowledge of other minds is in its turn dependent on knowledge of the objective world, because triangulation in addition requires that speaker and interpreter recognise that they share the world. To Davidson, the objective world is an intersubjective, common world.

Davidson thus defends an externalism that mixes causal and social considerations (1991b, 200f). Social factors connect with perceptual externalism in that the second person is necessary for locating the speaker's object of thought. If there is no way an interpreter can

match the speaker's utterances with objects in the world, we are not entitled to say that the uttered sounds belong to a language.

In his later articles, Davidson embraces the view that linguistic meaning is socially constituted, that is, necessarily intersubjective. Aspects of causes determine the content of thoughts, and which these causes are in the particular cases can only be established socially, by triangulation. This means that causes merely constitute content when socially identified (Davidson 1991b, 202; *cf*. 1991a, 165). Thus, via triangulation social factors play a constitutive role for meaning.[5]

3. The move from perceptual to linguistic discrimination

Davidson's claim that without a second person, utterances would not have meaning has received a lot of criticism. C. J. L. Talmage (1997) criticises Davidson from an individualist position, which recognises that linguistic meaning must be public, while denying that is must be socially constituted. As J. Fennell (2000) argues, Davidson may be interpreted as holding the weaker view that meaning necessarily is public. On this view, meaning still can be individual. It only excludes private meaning. Fennell furthermore notes that in later articles, Davidson emphasises the stronger view that meaning is socially constituted.

---

[5] W. Child (2001) argues that Davidson puts forward a modest realist account, in agreement with crucial parts of H. Putnam's realism about natural kinds, as presented in "The meaning of 'meaning'" (1975). As Child himself observes (2001, 29), Davidson accentuated certain elements of his theory towards the end of the 80ies and onwards, elements that made Davidson refrain from describing himself as a realist. I find it hard, if not impossible, to see how the view that Davidson expressed in his later work might be reconciled with Putnam's. Compare Davidson's view with the following account of Putnam's theory, which relies on his "The meaning of 'meaning'". Putnam holds that the extension of terms for natural kinds is fixed by ostension. An ostensive definition introduces a particular term. Every subsequent sample that bears the theoretical relation of being the same essence (*i.e*. hidden structure) as the sample used in the ostension will be referred to by that term. Terms for natural kinds are rigid designators: they designate the same kind of entity in every possible world. It is the microstructure of the kind that determines the reference of the term. Putnam writes that "'we use the term rigidly' to refer to whatever things share the *nature* that things satisfying the description normally possess" (1975, 238). A stereotype is normally associated to such terms. It contains information about defeasible recognitional criteria, *i.e*., features that in normal situations constitute ways of recognizing if a thing belongs to the kind. Putnam leaves it open whether we should say that the extension or the criteria constitutes the meaning of the term. He writes that "the extension of our terms depends upon the actual nature of the particular things that serve as paradigms [samples], and this actual nature is not, in general, fully known to the speaker" (1975, 245). The criteria are constrained by our perceptual apparatus and determined socially, while indexicality, or extension, is determined by the environment, independently of how speakers react to it. Davidson does not make any similar distinction between recognitional criteria and extension. He argues that content is constituted by concepts that are socially individuated, also in the case of triangulation. Thus Putnam and Davidson do not agree on how the environment contributes to language. Child maintains that the key idea of Putnam's account is that "explaining the meaning of a term involves the use of paradigms" (2001, 42). But in the present context the crucial difference between Putnam and Davidson lies in how they conceive of the objects of reference. Putnam is a full-blooded realist. Davidson, who claims that the objective world is the intersubjectively constituted world, holds that the objects of reference are those aspects of a cause that are socially determined. It is not possible to directly refer to, or talk about, essences of kinds. For instance, Davidson writes that "(C)ommunication, and the knowledge of other minds that it presupposes, is the basis of our concept of objectivity, our recognition of a distinction between false and true belief. There is no going outside this standard to check whether we have things right", and further down on the same page "(A) community of minds is the basis of knowledge; it provides the measure of all things. It makes no sense to question the adequacy of this measure" (1991a, 164). Davidson could not accept Putnam's 'semantic' theory of extension, while he, of course, might agree with Putnam's 'pragmatic' theory of how terms are used.

Talmage asserts that Davidson does not succeed in establishing that a speaker's utterances must be not only interpretable, but in fact interpreted, to be meaningful (1997, 139). She holds that what a speaker means is determined by how the speaker herself conceives of the stimulus of her utterance. An interpreter, or second person, is not necessary to determine the content of the utterance.[6] I will argue for the stronger view that meaning is socially constituted, against Talmage, by distinguishing between two kinds of discrimination of sensory information: perceptual and linguistic.

Talmage attacks what she takes to be Davidson's major argument in support of the view that triangulation is necessary for locating the object of thought. She first argues against the claim that triangulation depends on that speaker and interpreter share innate similarity responses. She declares that given that it is correct to say with Davidson that the speaker's innate similarity responses are to objects in the world, and not to patterns of stimulation, it does not seem necessary that an interpreter fix the content (1997, 142). On the other hand, she continues, if the similarity responses on the contrary are not shared between speaker and interpreter, there is not any point in appealing to the interpreter. Either way, interaction based in similarity responses cannot be essential for locating the object, she claims.

Talmage goes on to examine the view that a concept of the objective world is necessary for having a concept of an object. She concedes Davidson's claim that a second person is necessary for acquiring the notion of an objective, or intersubjective, world (1997, 144). She then declares that nevertheless, it does not follow from this that an interpreter is necessary for either the speaker's utterance to be meaningful, or the speaker to have concepts of entities in the objective world.

I believe, contrary to what Talmage argues, that the second person is necessary for thoughts to have objective, or intersubjective, content. Yet, thoughts may have content also when uninterpreted. But the content that they then have is not objective in the required sense. It is individual and based in perception. We should distinguish between the individual's perceptual recognitional capacities, with a basis in human (psychophysical) constitution, and those recognitional capacities that are anchored in (social) linguistic practise. Perceptual and linguistic discrimination impose different categories on information.

Linguistic discrimination makes use of linguistic concepts that are linked to natural languages. Linguistic concepts are symbols, that is, signs that are used to stand for something else that does not necessarily have to be present in the actual world, but may be fictive or abstract. All symbols are detached from their referents, and some are independent of the

---

[6] P. Pagin (2001) also denies that triangulation is necessary for bestowing thoughts or utterances with a content. He holds a weaker thesis concerning the role of triangulation. He maintains that triangulation is needed for the intentional explanation of behaviour. Pagin suggests that in explanations of behaviour we identify the contents of other speakers' thoughts by the help of a general principle. The principle concerns inference to the best explanation, and in this particular case, of the success of communication. Pagin claims that the best explanation of such success would be that speakers communicate about objects and events that are easily identifiable by being (perceptually) salient. On these grounds, Pagin maintains, we have reason to hold that whenever two speakers communicate, they do so about something they both observe, and moreover believe they both observe. What they both observe is located in the intersection of the causal lines between the speakers and the object, as in triangulation. In section 4, I argue that objects (or other perceptually identifiable entities) in fact rarely are easy to identify simply by saliency. The saliency of an object is a function not only of its properties, but also of the context in which it appears, and of the subject that apprehends it. As a result, a certain object may be salient to one speaker, but not to another. The object that the interpreter takes to be salient may not coincide with the object that the speaker takes to be salient. Furthermore, there may not even be any salient entities in the context of utterance. Consequently, it will not do to assume as a guiding principle in explanations of behaviour merely that thoughts are about salient entities.

spatiotemporal existence of their referents (Brinck and Gärdenfors 1999). Tomasello (1998) further describes symbols as social, intersubjective, and reciprocal, or bi-directional. That symbols are reciprocal means that a competent speaker understands that different speakers can use the same symbol with the same intent. If a speaker understands a symbol, she can also produce it, and *vice versa*.

Linguistic concepts are interrelated semantically by, for instance, opposition and substitutability, and can be defined in terms of each other. They are as well inferentially related, by operations that rely on stable core meanings or categories that enter into propositions. Stable core meanings and categories are important for the generality, or context-independence, that underlies the systematic and productive character of linguistic concepts and that results in linguistic creativity.

Finally, linguistic discrimination relates concepts within a framework that determines how existing concepts should be applied when speakers stand before new information. It also suggests how the concepts might be extended, and which new concepts may be introduced while keeping the conceptual network coherent. This makes it possible for speakers to agree on how to interpret new data.

Perceptual discrimination, on the other hand, is context-sensitive and tuned to the subject. Perceptual categories are based in the innate properties of individuals, and develop when the individual interacts with her environment. Small infants initially use spatiotemporal information to identify objects. Around one year they also start using feature-based information (Xu and Carey 1996). The perceptual categories that children develop are causally related to the objects that the categories represent. These categories characterise objects by their functional or perceptually accessible, surface properties (Samuelson and Smith 1999; Samuelson and Smith 2000).[7] Thus perceptual categories have a pragmatic nature and are based in the individual's physical interaction with the world. In the adult, perceptual categories and linguistic concepts exist side by side.

Talmage seems to neglect the difference between perceptual and linguistic discrimination. She claims, for instance, that given that it is correct to say with Davidson that the speaker's innate similarity responses are to objects in the world, it is not necessary that an interpreter fix the content. But once we realise that the similarity responses are to perceptually categorised, idiosyncratic objects, and not to entities that correspond to linguistic concepts, the importance of the second person for the fixation of content becomes clear.

The role of triangulation is to make it possible for speaker and hearer to agree on a common categorisation of the cause by jointly focusing on one of its aspects. The second person is necessary for moving from individual perceptual discrimination to shared linguistic discrimination, and thereby for conceptually identifying those objects in common space which answer to the speaker's utterances. The way in which the speaker herself perceptually categorises the stimuli is not inherently, nor constitutively, connected to the concepts that belong to linguistic discrimination.[8]

---

[7] The functional properties of an object are those that the object acquires in use. They can be described in terms of what the object affords to the subject. Affordances define the potential for action for a particular individual or a particular species in a given situation or environment, and have a physical reality (Zaff 1995, 239). They define the sum total of measurable actions that are possible by a given individual within a given context.

[8] Note that the present view does not exclude that linguistic concepts are indirectly grounded in perceptual categories. However, it does exclude the view that semantic content emerges directly from individual perceptual experience.

4. A problem for triangulation

Davidson asserts that knowledge about other minds, as well as about a shared world, emerges holistically and is inherently interpersonal. We stand before an interdependence between three kinds of knowledge: of one's own mind, of other minds, and of an objective, intersubjective world. None is conceptually, nor temporally prior, but communication depends on all three kinds.

In spite of the interdependence of the three varieties of knowledge, Davidson's theory of interpretation seems at first to pull in opposite directions, one conceptual and meta-representational, the other causal and behavioural. For instance, interpretation requires having higher-order thoughts about one's own as well as the other speaker's mental states. Triangulation, on the other hand, depends on making similarity judgements about behaviour in order to perceptually zoom in on a common cause. It does not necessarily involve representing the intentional states of the other person. At one point, Davidson describes the social interaction of triangulation as pre-linguistic and pre-cognitive (1997b, 129).

Speaker and interpreter converge on a common cause by observing each other's responses to external situations. Their responses tend to be similar, as a result of evolution and innate learning mechanisms. Nevertheless, the way in which the common cause is determined cannot be merely behaviourally based, but is also based in inferences about intentional states. Although the beliefs about the similarities of responses are externally caused, the similarity judgements that the speakers as interpreters make about each other are about their respective ways of conceiving of the world, and not just about behaviour. Behaviour simply provides the impetus for making inferences about the beliefs of the other person. After all, knowledge of other minds is, according to Davidson, necessary for triangulation.

As mentioned in section 1, the role of triangulation is to break the intentional circle by providing non-semantic evidence for what words mean. But if triangulation depends on the ascription of beliefs to oneself and other speakers, we are back to square one. There are no clean-cut causes from which communication can take its beginning. We have reached a dilemma. Either triangulation is pre-cognitive and cannot explain how speakers converge on a common cause, or it requires higher-order thought, and cannot provide for language entry, nor individuate the content of individual beliefs and utterances.[9] This is where, in the theory of radical interpretation, charity shows up again, especially in the guise of the principle of correspondence. Triangulation is supposed to help getting interpretation going and supplement charity. Likewise, charity is supposed to back up triangulation, by guaranteeing that during triangulation the speakers' similarity judgements correspond. — Unfortunately, it cannot do so. Here is why not.

While Davidson highlights the use of higher-order intentions during interpretation, he thinks that the convergence of similarity judgements depends on quite basic and simple mechanisms. Due to evolution and conditioning, natural habits and generalisations correspond across human beings (Davidson 1991a, 212; Davidson 1991b, 202). Davidson believes that sample ostensions and a natural knack of induction make the transition from the pre-propositional to the propositional sphere effortless (1999, 42). This avowal seems overly

[9] J. Fennell (2000, 150) essentially points to the same tension as I do. However, Fennell argues that since linguistic meaning is socially constituted, the dilemma cannot be resolved. As I will show, this conclusion does not follow. It is possible to provide for language entry without giving up the social account of meaning.

optimistic. The gap between sensory perception and language cannot, I maintain, be bridged that easily. It is no doubt correct that triangulation depends on cognitive similarities between speakers, some of which are socially grounded, some of which have their origin in human evolution. But these traits will seldom do to isolate the common cause in the way that is required for concept attribution.

One fundamental problem for triangulation is that the principle of correspondence, which lends support to triangulation, concerns only the general, or shared, properties of speakers, while triangulation needs support in the individual cases of interpretation. To be successful, triangulation requires that contextual properties be taken into account in order to determine the common cause. The individual cases are hence central to radical interpretation. To explain why and how they are central, let me return to the process of triangulation.

When triangulating, the interpreter groups together the speaker's responses to some feature in the context of utterance on the basis of the innate abilities that she shares with the speaker. Having grouped together a number of the speaker's responses, she will then be able to attribute the concept of an *x* to the speaker. Responses are put in the same group on the grounds of being similar.

But, I submit, it is not responses taken as mere behaviours that are grouped together, but responses taken as goal-directed, intentional actions that are performed in response to some particular feature of the context. Intentional actions cannot be identified by their behavioural properties alone, taken in isolation from the feature that triggered them. For instance, the same gesture, say, a repeated waving of the hand, may in one context be a way of saying "hello" to somebody else, while in another context the gesture simply serves to cool oneself. The gesture has the same structure, but different functions in the two situations.[10]

Thus for the interpreter's categorisation of the speaker's responses as similar with one another to be of any help in interpreting, the interpreter must conceive of the responses as at least being to the same kind of feature. If not, there will not be any reason for her to suppose that the speaker's responses should form an equivalence group with her own responses, even if the responses were similar from a purely behavioural point of view. And then there would not be any reason to take the responses as evidence for attributing a concept of an *x* to the speaker.

But, you may object, in the long run triangulation presumably will isolate the feature to which the speaker responds. Repeated observations of the speaker's behaviour in similar contexts will probably converge on the factors that are central, while sifting out what is irrelevant. Eventually triangulation will succeed in isolating the feature that triggers a certain kind of behaviour. Why would this not be sufficient for the success of interpretation? — Because to at all be able to make such repeated observations, the interpreter must hold something constant across contexts of observation. There must be some way of telling that the observations are of a similar kind of behaviour. The strategy is comparable to the one taken by the radical interpreter, who in order to break into the intentional circle holds belief constant, while solving for meaning. However, in this case the interpreter holds the triggering feature constant.

---

[10] The structure of a behaviour concerns its components (subparts) and the way these are ordered, while the function concerns its use. For instance, pointing can be done with the arm, the whole hand, or the index finger, and be accompanied by looks to a partner before, during, or after the gesture is made. The differences reflect different structures of the behaviour (Leavens and Hopkins 1999). Further, pointing can have different functions, *e.g.* to indicate and share attention to an object, or to urge the partner to get an object.

Furthermore, the feature that triggers the speaker's behaviour, and the observation of which will help distinguishing that particular behaviour from other behaviours that are structurally, but not functionally, similar, is not identical to the common cause. The common cause is identified by repeated observations.

Remember that a common cause in Davidson's terminology corresponds to an aspect of an object. Concept attribution depends on the possibility of determining common causes. However, isolating and keeping constant the triggering feature across contexts of utterance precedes the conceptual identification of a common cause. This identification is achieved only in the long run. The contextual feature that is held constant across contexts will be more coarsely individuated than an aspect of a cause is. The contextual feature is not conceptually individuated, and cannot be so, until interpretation is well on its way. During the initial phase of radical interpretation, the interpreter can access only less fine-grained elements of the context of utterance.

In view of what has been said, it seems that any judgement about responses across contexts of use will depend on previous judgements about the triggering feature of single responses. Triangulation presupposes that it is possible to discern which feature the speaker reacts to in advance of interpretation. Moreover, it could be argued that the interpreter will only be justified in grouping particular behaviours together, if she has reason to believe that what she takes to be the triggering feature of the speaker's response corresponds to that which the speaker herself takes to be the triggering feature.

Hence, the interpreter appears to have two options. She might simply assume that her own categorisation of the sensory information is naturally similar to the speaker's. She might instead perceptually identify the source of the sensory stimuli, which prompts the speaker's utterance, in a similar manner as to how the speaker identifies it. This should be done independently of the speaker's own conceptual framework. Yet, the possibility of in this way identifying the source in advance of the actual interpretation is, of course, denied by the theory of radical interpretation.

But could not the cognitive correspondence between speaker and interpreter explain how the interpreter's categorisation of similarity judgements might succeed? Certainly, it is safe to assume with Davidson that humans roughly tend to make similarity judgements in unison. But that they do so, however, does not ensure convergence in particular situations. And the individual situations are crucial to get radical interpretation off the ground. Likewise, although agreeing upon similar principles of rationality, as well as being of the same species, of the same culture, or even of the same (human) family, may all be necessary for the convergence of similarity judgements, these facts are far from sufficient to reliably isolate the triggering feature of a particular response.[11]

Conditioning, because it is a learning process that concerns particular individuals and situations, may explain how the triggering feature can be discerned — provided that speaker and interpreter have been conditioned to similar situations. But even then, there is a great deal of uncertainty in particular contexts of interpretation. Being conditioned to react in certain ways to sensory stimuli is a trait that is constant across contexts for the speaker. It alone cannot guarantee the interpreter's success in determining the individual utterances' triggering

[11] E. Picardi (1999, 180) argues that Davidson's notions of shared similarity and shared interests only make sense as against a background of socially or culturally shared practices. This claim may be correct, but it cannot alone explain how similarity judgements happen to converge in the individual case. It fails for the same reason as Davidson's background assumptions about cognitive similarity do: it is too general.

feature. The process of determining the nature of incoming sensory stimuli is always tuned to the context and the speaker's internal state at that time.

Davidson (1999) suggests that perceptual saliency provides the grounds on which the interpreter safely can assume that the speaker reacts to those objects that she herself discerns. Indeed, humans do naturally find some features more salient than others, most certainly because of evolution, innate learning mechanisms, and social conditioning. Moreover, saliency is a context-sensitive feature that depends on properties of the context that are accessible to both speaker and interpreter. Yet, saliency does not remove the indeterminacy that occurs in the context of triangulation, for the following reasons.

Perceptual saliency is connected to how much effort it takes to perceive an item. The most salient features are the ones that can be picked up with the least effort. Usually these are the ones that contrast against our expectations. However, expectations are not always shared among speakers, but can be highly individual. While sharing a context or situation, different speakers may find different features of it salient. For instance, different speakers may be more or less, or perhaps not at all, familiar with the context. They may as well have different backgrounds, which will influence their expectations on the context, and thereby their perception of it. Each speaker has an individual history that is forming her perception and her expectations on what she actually will perceive in a particular context.

An additional complication is that the interpreter may not know which difference or contrast to look for in order to pick out the triggering feature. For one thing, there may not be any particularly salient item *per se*. There may be several competing items in the context, or no especially salient items present at all. Then the interpreter needs something to guide her search for saliency: she needs to know which contrast is relevant on that occasion. But in this kind of situation, relevance is connected to the aim that the speaker has with her utterance. To determine what is relevant, the interpreter will have to engage in interpretation of the speaker's intentional states. However, this involves ascribing thoughts with contents to the speaker — contents that the interpreter in principle cannot ascribe at this point, according to the theory of radical interpretation.

5. Finding an entering wedge

Davidson has two separate motives for putting forward his theory of triangulation. Broadly speaking, Davidson's theory deals with the problem of the location of the stimuli from which meaning emerges. To give a location to the cause of a thought is to define its content (Davidson 1991a, 159).[12] But the problem is ambiguous. It concerns, for one thing, the metaphysics of meaning, or what a theory of interpretation is about (*cf*. Heal, 1997). It also concerns the individuation of content, that is, how to apply a theory of interpretation to particular utterances (*cf*. Picardi 1999).

As a theory of the first kind, which contributes to a clarification of the nature of meaning, it discards the need to postulate 'meaning entities', such as sense-data, propositions, or similar kinds of objects of thought (Davidson 1973, 126; Davidson 1987, 35; Davidson 1998). It aims to reject views that conceive of meaning as an object before the mind's eye, by,

[12] Until the triangle is complete, there can be no answer to the question whether a response is to proximal stimuli at the sensory surfaces, to distal stimuli, or to stimuli located somewhere else along the link from the cause to the response it induces (1991a, 159). As noted above, Davidson holds that the response is not to the cause as such, but to an aspect of it.

as I see it, successfully providing an account of how utterances can have meaning that does not presuppose such entities. The meaning of a sentence depends on its logical relations to other sentences, as well as on its evidential relations to other sentences and the objective world.

As a theory about the individuation of content, Davidson's theory is less satisfactory. As argued in section 4, it cannot account for how particular utterances receive an interpretation. Triangulation may give non-semantic evidence for the meaning of utterances, but in the end, the common cause cannot be isolated without resorting to a linguistic interpretation of the speaker's beliefs. Although triangulation relies on causal links between the speakers and their environment, it also requires having beliefs about one's own and the other speakers' mental states. Moreover, it requires having the concepts of error and of the subjective and objective realms. The intentional circle remains intact. Davidson writes:

> The interdependence of thought and language poses a difficulty for theory (…) for it denies us the chance of finding an entering wedge. As a result, we can neither say in detail how the transition from the pre-propositional to the propositional takes place, nor can we reduce the propositional to the pre-propositional (1999, 42).

Depending on whether we work with a weak (non-intentional and pre-cognitive) or strong (intentional and propositional) notion of triangulation, triangulation either cannot improve on charity, or will not be of any help during the first phase of interpretation, language entry.

I suggest that we approach the problem of how to get interpretation started by thinking of triangulation in terms of joint attention.[13] This approach will be explored below. Joint attention consists in the ability of two or more subjects to focus their perception simultaneously, as a consequence of attending to each other, on a shared object of attention. Joint attention has the advantage of not being a purely causal or behavioural process, while all the same not necessarily standing in need of the speakers' having intentional states about intentional states (Brinck 2001).

In section 4, I argued that for triangulation to provide for language entry, it must initially isolate the triggering feature of the speaker's response in the context of utterance. Yet, the cause cannot be identified conceptually, since in the case we are considering, radical interpretation starts from scratch. On the other hand, to identify it purely causally is useless, since it then could not serve to bridge the pre-propositional and propositional spheres.

In the remainder of this article, I will present an outline of a theory of joint attention, according to which attention is jointly focused on objects that are identified both perceptually and intersubjectively, or socially (I use, as Davidson does, the latter words interchangeably).

---

[13] K. Bühler (1934) stressed the importance of joint attention for determining the meaning of utterances. Bühler maintained that utterances take place in two contexts simultaneously: the symbolic and the deictic fields. The utterances involve both symbols and deictic words. The meaning of symbols is determined by a co-ordination with objects and states of affairs in the external world. The meaning of deictic words is incomplete. It is made definite from case to case by sensory information provided by the deictic field, which is centred on the speaker. Bühler focused on the role of joint attention in determining the meaning of utterances for competent speakers of the same language. In Bühler's so-called organon model, attention is primarily directed at the semantic and acoustic properties of the linguistic sign, properties that are known to all competent speakers. The case is slightly different for deictic words, since their meaning is completed in the context of utterance. But deictic words nevertheless have linguistic properties, which are known to all competent speakers. Consequently, the focus of Bühler's model of joint attention is different from the focus of the model that will be presented here. The role of the model put forward in this article is to explain how language entry is possible. It concerns a situation in which speaker and interpreter do not share a language and thus are not familiar with the linguistic properties of their respective languages. This means that the present model deals with joint attention as a process that may determine the content of an utterance independently of a shared language (or of a shared theory of interpretation).

These objects will eventually, as interpretation proceeds, be conceptually individuated. The theory is designed to show that it is possible to break into the intentional circle without giving up either the irreducibility of linguistic meaning, or the interdependence of propositional belief and linguistic meaning. The suggestion that joint attention may serve to break into the intentional circle is not supposed to be exhaustive. There may be other ways to do so, which do not involve joint attention.

In introducing the notion of joint attention, we have the choice between two claims concerning triangulation. We can hold that there is only one kind of triangulation, that is, joint attention, as described above. Or we can hold that there are two kinds, one weaker, joint attention, and one stronger, which I will call semantic triangulation, following Pagin (2001). I believe that the latter option, of having two kinds of triangulation, is preferable. I agree with Davidson that the concept of meaning is irreducible, and that linguistic meaning cannot be naturalised. What a linguistic expression means must be determined from within the language. Consequently, semantic triangulation is necessary for the interpretation of natural languages. However, that linguistic meaning is irreducible does not imply that intentional communication as such necessarily is linguistic. Nor does it imply that joint attention is irrelevant to linguistic interpretation. Joint attention can pave the way for language entry and get interpretation going.

6. Joint attention

For present purposes, the requirement on joint attention is that the process should issue in a perceptual discrimination of a common cause, a discrimination that could serve as the basis for an ascription of propositional content. This discrimination should be possible without engaging in an explicit interpretation of the intentional states of the speaker. The aim of this section is to present an analysis of the act of joint attention that obeys these constraints, and which furthermore agrees with empirical research concerning joint attention.

Suppose that a speaker utters a sound with the intention to direct the attention of the hearer to a particular object in the surroundings. Suppose furthermore that the sound draws the attention of the hearer to the speaker. Then it seems that we can spell out the conditions for an act of joint attention as follows. Joint attention will occur iff speaker and hearer

(i) attend to each other's states of attention,

(ii) make attention contact, and

(iii) alternate gaze between each other and the object.

Notice that the ordering of (i)-(iii) is not temporal.

The first condition for joint attention concerns the way in which the speakers must attend to each other in order to engage in the process that eventually will result in joint attention. Several authors have refined this condition. They emphasise that joint attention requires not only that the speakers attend to each other's states of attention, but that they attend to those states in a certain manner (Bruner 1998; Gómez 1998; Tomasello 1998). The speakers should grasp that attentional states are directed at objects (or events, other beings, *et cetera*) in the environment, and by being so directed are about them.[14] This claim underlines that in the process of joint attention, the speakers do not only attend to each others' behaviour,

[14] The formulation in the text above reminds of Brentano's analysis of intentionality. The close connection between attention and intention has been stressed by researchers in different fields (*e.g.* Eilan 1998 and Morrison 2001).

but indirectly to the intention behind the behaviour (yet without entertaining higher-order thoughts about it). It expresses the fact that to be able to achieve joint attention, the speakers must have grasped the difference between mere behaviour and intentional action. They have to understand that intentional action is directed at an object and has a purpose.[15]

But the first condition for joint attention cannot in its present formulation explain how the process of joint attention can begin in the absence of salient entities. It stands in need of a final refinement (Brinck 2001; *cf.* Emery et al. 1997). As noted in section 4, in many cases there is not any particularly salient entity present in the context of utterance. Accordingly, the speakers should attend to each other's states of attention also when these states are directed at non-salient entities. Unless this condition holds, the hearer will only follow the attention of the speaker in the presence of a (to the hearer, at least) salient object. But communication can, and sometimes does, succeed in the absence of such objects. Attention-getting behaviour will trigger gaze following also when the hearer's attention is not attracted by any salient object. Then attention focusing is guided solely by the speakers' mutual attention to each other. Achieving joint attention does not necessarily depend on that the environment provides information about which object is relevant.

Let me illustrate my point. Suppose that the speaker has seen a third person hide something in a box. She wants to indicate the box to the hearer. The box stands on the floor among several other boxes. All the boxes look alike. There is no evidence that one of them contains something of value. In this case, the speaker would only be able to make the hearer take part in the process of joint attention if the following were the case. The hearer should be sensitive to the speaker's manifest interest in one of the boxes. She should moreover react to the speaker's interest as if it were a motive for engaging in attention contact with her and for following the line of her gaze to the box.

Hence, we can reformulate the first condition for joint attention as follows. Joint attention will occur iff speaker and hearer

> (i') attend to each other's states of attention as states that are directed at objects in the environment and by being so directed are about the objects, and, moreover, attend to these states also when they are directed at non-salient entities.

The reader may object that joint attention can occur without conditions (i')-(iii) being fulfilled. It is true that there are many ways in which agents can focus their attention on a shared object, some of which do not require the concepts of the other as an intentional agent, of a common world, of a shared object, *et cetera*, and do not involve gaze alternation, nor attention contact.[16] But not all kinds of 'joint attention' obey the constraints presented in the beginning of this section. Furthermore, our concern is with the kind of joint attention that allows for intentional communication and paves the way for language entry, and the conditions are meant to cover that kind.

---

[15] Children are prepared to develop the capacity to perceive the intentionality of action during the first year (*i.e.*, they will develop it during normal circumstances). Already by the age of 14 months, children can imitate an action modelled to them by an adult. Imitation seems to require that the situational constraints of the adult are similar to those of the infant. It also seems that the child imitates the action that is seen by her or him to be the most rational way to reach the goal that is available within the constraints of the situation. Rationality is evaluated as a function of the visible goal of the action and the physical constraints of the actor's situation (constraints such as the presence of obstacles, the visual conditions, whether the adult's hands are blocked, *etc.*). Novel goal-directed actions are not automatically imitated, but are first evaluated as to their rationality (Gergely, Bekkering and Király 2001).

[16] Different kinds of joint attention are discussed in Brinck (2001) and Brinck and Gärdenfors (2003).

Now, if communication is to be successful, speaker and hearer must intentionally and consciously focus their attention on the same object or event, that is, on a common cause of stimuli. As emphasised by Davidson, common causes can never be constituted by the sense impressions of individual speakers or by proximal stimuli, but must occur in a common world, that is, a world that is shared by the speakers. Only causes that are publicly accessible can constitute the bases of meaning. On the other hand, the joint focusing of attention on the common cause should not be dependent on linguistic discrimination, or joint attention will not serve its role as an entering wedge.

In view of this, it seems reasonable to say that speaker and hearer will be able to jointly focus their attention on the common cause of some stimuli when, in a particular context of utterance (containing a time, a place, a speaker, and a hearer), the cause is mutually accessible to speaker and hearer by perceptual means.

Mutual perceptual accessibility is a necessary condition for joint attention, but far from sufficient. In addition, it is necessary to account for how speaker and hearer manage to perceptually identify the cause of the stimuli in a similar way and come to focus on a common cause. The second and third conditions for joint attention stated above explain this in terms of attention contact and gaze alternation.

Attention contact occurs when a speaker is attending to the attention of another speaker, who in turn is attending to the first speaker's attention (Gómez 1994, 72f). This usually (and crucially, for present purposes) involves making eye contact, and thus noticing, by perceptual means, that the other is attending to you and to where you yourself are directing your (visual) attention.[17] By making attention contact, the speakers can direct each other's attention by gaze, or gestures, or both. Directing the attention of another speaker involves moving one's gaze back and forth between the object and the speaker, as well as a repeated visual checking of the target of the other speaker's attention (as the attention manifests itself in behaviour). By engaging in attention contact and guiding each other's attention through gaze alternation, the speakers can co-ordinate their perception, and intentionally and consciously jointly focus on a common cause.

Joint attention is supposed to lead to language entry. Ideally, the speaker uses a perceptual strategy to direct the attention of the hearer onto an object, while the hearer identifies the object by linguistic means as soon as she has focused her perceptual attention on it. But a correct interpretation and ascription of meaning to the speaker's utterance will in principle not be possible until the radical interpreter has a theory for the language as a whole. Moreover, since radical interpretation, according to Davidson, is indeterminate, that theory may just be one of several ways of making sense of the evidence.

Obviously joint attention is not a foolproof way of identifying the common cause.[18] The evidence that makes the interpreter assign a certain content to the speaker's utterance is not conclusive. Radical interpretation still is operative. Nevertheless, joint attention

---

[17] Eye contact plays an important role for joint attention (*cf.* Baron-Cohen 1995). Individuals born blind eventually seem able to compensate for the lack of eye contact, but it appears that the impairment delays, or in some cases even damages, their capacity for solving tasks that require having a properly functioning theory of mind (*cf.* Garfield, Peterson and Perry 2001, 512f).

[18] The speakers' interests, needs, and motives, as well as the value of the object, and the emotions the object evokes in the speakers are all factors that may play a positive role for achieving joint attention. An analysis of their role for joint attention lies beyond the scope of this article.

constitutes a way of breaking into the intentional circle, precisely because it is intentional without being propositional. In the next section, I will explain how the circle can be broken.

7. Understanding speakers' intentions

I have claimed that neither joint attention, nor the process that leads to it, are propositional, and furthermore, that they do not require having intentional states about intentional states. But how then do the speakers express and understand the communicative intentions that are involved in the process?

Suppose that the speaker intends to indicate an object to the hearer. The context is one in which speaker and hearer do not share a language. In order to communicate the content of her utterance, the speaker attracts the attention of the hearer by the sound of her utterance, perhaps together with a gaze, a pointing gesture, or both.[19] She then attempts to direct the attention of the hearer to the object by alternating her gaze between the object and the hearer. She intends to make the hearer focus her attention on the target. To respond appropriately, the hearer must recognise that the speaker intends to communicate, and moreover intends for the hearer to focus her attention on a target.

How can non-propositional, or non-linguistic, communication be intentional in the manner just described? In order to answer this question, I will first consider Grice's analysis of non-natural meaning. Then I will present a version of that analysis that shows how it is possible to recognise speakers' intentions without engaging in higher-order thoughts about the intentions of either oneself or other speakers.

According to Grice's analysis of non-natural meaning, the speaker has beliefs about the capacity of her utterance to produce certain beliefs (or other intentional states) in the hearer (Grice 1957). It is the fact that the speaker intends the hearer to recognise her underlying intention to produce a certain belief in the hearer that distinguishes non-natural from natural meaning (Grice 1957). The speaker is aware of the informational effect of her message, and her being so involves both self- and other-directed intentions.

Gricean models of intentional communication (such that conveys non-natural meaning) usually contain intentional loops that involve speaker and hearer. The speaker intends (i) her utterance to produce a certain (cognitive) effect in the hearer, (ii) that the hearer recognises the speaker's intention (i), and (iii) that the hearer's recognition of the speaker's intention (i) shall function as at least a part of the hearer's reason for her response to the utterance (*cf*. Grice 1957). But, as presently will be argued, intentional loops are not necessary to describe how either speaker or hearer can be aware of the informational effect of a message — although we are dealing with intentional communication that conveys non-natural meaning. When the message is imparted by a process aiming at shared reference by joint attention, manifest attentional states will do the job.

According to Grice, a message conveys non-natural meaning when the hearer's response to it results from his or her recognition of the speaker's intention with the utterance. The hearer's recognition of the underlying intention should not be causally induced, but be within the control of the hearer (Grice 1957). It should be a reason or motive that moves the

[19] In my analysis of the process that leads to joint attention, I do not discuss the role of vocalisation for that process. No doubt vocalisation is important.

hearer to recognise the intention, or the intention will be conveyed naturally, that is, be causally induced.

I agree that the hearer's response should not be merely causally induced. But to achieve joint attention, the hearer does not need to engage in higher-order thoughts about intentions to produce a response that is within her control (at least not during normal conditions). It is sufficient that she responds as a result of attending to the speaker's state of attention in the manner that was described by condition (i') in section 6. Thus the hearer's recognition of the speaker's attentional state, and not her intentional state, will provide the motive for the hearer's intention to attend to the object to which the speaker attends. On Grice's analysis of non-natural meaning, the underlying intention of the utterance is to produce a certain intentional state in the hearer. On the present analysis, the underlying intention of the utterance is to produce a certain attentional state in the hearer.

Suppose that the speaker wants the hearer to focus her attention on a particular object. The speaker imparts her message by engaging in attention-getting behaviour, making attention contact, and alternating her gaze between the object and the hearer's gaze. The speaker intends for the hearer to attend to her manifest states of attention. The speaker further intends for the hearer to attend to the object as a consequence of attending to the speaker's attention. In the normal case, the hearer will grasp the speaker's intention by perceiving the speaker's state of attention and then following her gaze to the object.

Why is it not enough that the hearer simply attend to the object as a consequence of attending to the speaker's attention-getting and attention-directing behaviour? Why should the hearer attend to the object as a consequence of having attended to the speaker's gaze by having attention contact with her? The reason why is that the hearer's engaging in attention contact and reacting to the speaker's gaze reflects her being implicitly aware of the speaker's communicative intention. The communicative intention is displayed by the speaker's attention behaviour, including her attempt to make attention contact with the hearer.

Thus we reach the following analysis of non-natural meaning as conveyed by acts of joint attention. The speaker intends

(i) her utterance to make the hearer attend to a particular object and share attention to it,

(ii) that the hearer attend to the speaker's attention in the correct manner (*i.e.* as described by condition (i')), and

(iii) that the hearer attend to the object as a consequence of having attended to the speaker's attention in the correct manner.

In principle, all the relevant information that is necessary for achieving joint attention is manifest in the immediate environment of the participants, and is accessible without engaging in a theory of mind. The speakers' states of attention reflect those intentions that are necessary to achieve joint attention to an object.

Only those states of attention that are manifest while the speaker is making the utterance and engaging in attention contact are relevant for the hearer. But not any state of attention that occurs during this process will be relevant for achieving joint attention. How will the hearer know which states are relevant? Suppose that a dog draws the speaker's attention while she is making attention contact with the hearer. The speaker looks at the dog, and then turns her gaze back to the hearer and resumes attention contact.

I submit that the hearer will be able to sort out those parts of the speaker's attention behaviour that are directed to her by their intersubjective character. The attention behaviour that is involved in attempts at joint attention is different from states of individual attention. During the process that leads to joint attention, the speakers actively seek attention contact and continuously interact during the process. They attend to each other's attentional states by

exchanging looks, gestures, utterances or sounds (as in non-linguistic vocalisation), facial expressions that show emotional reactions, and so on. In the example above, those of the speaker's states of attention that concern the dog do not have a social character. For instance, the speaker does not alternate her gaze between the dog and the hearer (*i.e.* she does not look back and forth between them).

J. C. Gómez (1994) has put forward a simplified model of intentional communication that is influenced by Grice's analysis. Gómez suggests that the use of the concept of attention contact provide a way to avoid multiplying higher-order intentions in the analysis of intentional communication. As mentioned in section 6, attention contact occurs when the speakers attend to each other's signs of attention by making eye contact and exchanging gazes. Propositional thought is not necessary for attention contact.

Gómez maintains that "eye contact implies mutual attention to each other's sign of attention" (1994, 73). I would like to stress that eye contact will only have this implication for subjects who are able to grasp that gaze is both intentional, that is, goal-directed, and displays interest. In a subject who does recognise this, the implication of eye contact for mutual attention will be immediate. Gaze thus acquires an imperative function. When the speaker, after having made attention contact with the hearer, ostensibly looks in some specific direction and subsequently alternates her gaze between the hearer and some object in the shared context, this directly urges the addressee to engage in joint attention by following the speaker's gaze. Consequently, there is a sense in which the signs of attention as well can be signs of communicative intentions.

Engaging in attention contact manifests an awareness of the role played by attentional states in intentional communication. It also constitutes a way of establishing something like the mutual awareness, or mutual knowledge, characteristic of the Gricean account, without having intentional states about intentional states (*cf.* Lewis 1969; Schiffer 1972). To avoid counterexamples to Grice's analysis, and explain how messages convey non-natural meaning, it has been suggested that most or all of the speaker's intentions with her utterance is known to the hearer, that the speaker knows that they are known to the hearer, and so on. This results in a, perhaps infinite, regress of intentional loops.

Making attention contact starts a process that constitutes an attentional loop that resembles the intentional loop. But in contrast to mutual knowledge, attention contact does not require having intentional states about intentional states. The awareness of the role of attentional states and attention contact for communication can be implicit: it does not have to be conceptually articulated or reflected upon. Yet it can be experienced, and the speaker can express his or her having this experience in action, by actively taking part in the joint attention process.

The ability to engage in attention contact and gaze following in order to focus on a joint target presupposes an understanding that attentional states may be intentionally directed (*cf.* the refinement of the first condition for joint attention). It requires that the hearer can grasp that attentional behaviour may communicate the speaker's intention, and thereby that the behaviour expresses a communicative intention. It also requires a more basic capacity, sometimes called attention reading.

Attention reading occurs when the hearer recognises that the speaker's attention behaviour expresses a particular kind of goal-directed intention towards an object in common space. It is exercised on the basis of observational and behavioural evidence gathered from, for instance, the facial expression, gaze, and body posture of the speaker. Attention reading constitutes the first step in the process that will lead to joint attention. By reading the speaker's attention and becoming interested in it, the hearer can join the speaker in this process.

As it happens, Davidson recognises a counterpart to attention reading at a propositional, or conceptual, level. He acknowledges the existence of a capacity of intention reading in speakers (1974; 1991a). He maintains that an interpreter from outward manifestations of propositional, intentional states can perceive which attitude a speaker has towards some object or event (1991a, 210). Intention reading makes it possible for the interpreter to detect the speaker's non-individuating, extensional attitudes to sentences, such as holding true, or preferring *p* to *q*, which do not distinguish between propositional contents. Davidson observes:

> The assumption that we can detect such an attitude does not beg the question of how we endow the attitudes with content, since a relation such as holding true between the speaker and an utterance is an extensional relation which can be known to hold without knowing what the sentence means (1991a, 210).

Intention reading resembles attention reading, but demands more of the speaker in terms of her cognitive capacities. In contrast to the intentional attitudes that Davidson describes, the attitudes that states of attention express are not directed at sentences or other linguistic entities. Neither do they require a grasp of the concept of truth.

Attention reading can, similarly to intention reading, be described as non-individuative. It takes more than grasping that the speaker is attending to a particular object to identify the object conceptually. As Davidson notes, for the object to be conceptually individuated, speaker and hearer must agree on a common conceptualisation of it. This is where joint attention and semantic triangulation connect. By reading the attention of the speaker, the hearer can go on to make attention contact, and engage in the process that leads to joint attention, and eventually to semantic triangulation. Joint attention to some aspect of a perceptually categorised object consequently prepares for concept attribution.

A theory of how to interpret behaviour by attributing perceptual, attentional states is very different from one, like Davidson's, that explicitly refers to propositional, intentional states. The latter kind of theory, which rationalises behaviour by attributing intentional states like beliefs and desires to the speakers, has a wider scope. But there are some things such a theory cannot do, such as explaining the transition from pre-cognitive behaviour to propositional thought, or how to begin to interpret a speaker of an unknown language. This is where a theory of joint attention has a role to play.

REFERENCES

Baron-Cohen, S. 1997, *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA.

Brinck, I. 2001, "Attention and the Evolution of Intentional Communication", *Pragmatics & Cognition*, 9(2), pp. 255-272.

Brinck, I. and Gärdenfors, P. 1999, "Representation and Self-Awareness in Intentional Agents", *Synthese*, 118, pp. 89-104.

Brinck, I. and Gärdenfors, P. 2003, "Co-operation and Communication in Apes and Humans", *Mind & Language*, 18(5), pp. 484-501.

Bruner, J. 1998, "Routes to Reference", *Pragmatics & Cognition*, 6(½), pp. 209-227.

Bühler, K. 1934 (1990), *Theory of Language. The Representational Function of Language*. John Benjamins, Amsterdam.

Child, W. 2001, "Triangulation: Davidson, Realism, and Natural Kinds", *Dialectica*, 55, pp. 29-49.

Davidson, D. 1973, "Radical Interpretation", *Dialectica*, 27, pp. 313-328.

Davidson, D. 1974, "Belief and the Basis of Meaning", *Synthese*, 27, pp. 309-323.

Davidson, D. 1975, "Thought and Talk", in: S. Guttenplan, ed., *Mind and Language*. Clarendon Press, Oxford, pp. 7-23.

Davidson, D. 1984, "Communication and Convention", *Synthese*, 59(1), pp. 3-17.

Davidson, D. 1986a, "A Coherence Theory of Truth and Knowledge", in: E. Lepore, ed., *Truth and Interpretation*. Blackwell, Oxford, pp. 307-319.

Davidson, D. 1986b, “A Nice Derangement of Epitaphs”, in: E. Lepore, ed., *Truth and Interpretation*. Blackwell, Oxford, pp. 433-446.
Davidson, D. 1987, “Knowing One’s Own Mind”, in: *Subjective, Intersubjective, Objective*. Oxford University Press, Oxford, 2001, pp. 15-38.
Davidson, D. 1988, “The Myth of the Subjective”, in: M. Krausz, ed., *Relativism. Interpretation and Confrontation*. University of Notre Dame Press, Notre Dame, pp.159-172.
Davidson, D. 1991a, “Three Varieties of Knowledge”, in: A. Phillips Griffiths, ed., *A. J. Ayer Memorial Essays. Royal Institute of Philosophy Supplement 30*. Cambridge University Press, Cambridge, pp. 153-166.
Davidson, D. 1991b, “Epistemology Externalized”, *Dialectica*, 45(2-3), pp. 191-202.
Davidson, D. 1992, “The Second Person”, in: *Subjective, Intersubjective, Objective*. Oxford University Press, Oxford, 2001, pp.107-122.
Davidson, D. 1997a, “Indeterminism and Antirealism”, in: *Subjective, Intersubjective, Objective*. Oxford University Press, Oxford, 2001, pp. 69-84.
Davidson, D. 1997b, “The Emergence of Thought”, in: *Subjective, Intersubjective, Objective*. Oxford University Press, Oxford, 2001, pp. 123-134.
Davidson, D. 1998, “The Irreducibility of the Concept of the Self”, in: *Subjective, Intersubjective, Objective*. Oxford University Press, Oxford, 2001, pp. 85-91.
Davidson, D. 1999, “Interpretation: Hard in Theory, Easy in Practise”, in: M. de Caro, ed., *Interpretations and Causes*. Kluwer, Dordrecht, pp. 31-44.
Eilan, N. 1998, “Perceptual Intentionality, Attention and Consciousness”, in: A. O’Hear, ed., *Contemporary Issues in the Philosophy of Mind*. Cambridge University Press, Cambridge, pp. 181-202.
Emery, N. J., Lorincz, E. N., Perret, D. I., Oram, M. W. and Baker, C. I. 1997, “Gaze Following and Joint Attention in Rhesus Monkeys (*Macaca mulatta*)”, *Journal of Comparative Psychology*, 111(3), pp. 286-293.
Fennell, J. 2000, “Davidson on Meaning Normativity: Public or Social”, *European Journal of Philosophy*, 8(2), pp. 139-154.
Garfield, J. L., Peterson, C. C. and Perry, T. 2001, “Social Cognition, Language Acquisition and the Development of the Theory of Mind”, *Mind & Language*, 16(5), pp. 494-541.
Gergely, G., Bekkering, H. and Király, I. 2001, “Rational Imitation of Goal-Directed Actions in 14-month-olds”, Poster presented at the *23rd Annual Conference of the Cognitive Science Society*, 1-4 August 2001, Edinburgh, Scotland, pp. 1-5.
Gómez, J. C. 1994, “Mutual Awareness in Primate Communication: A Gricean Approach”, in: S. T. Parker, R. W. Mitchell and M. L. Boccia, eds., *Self-Awareness in Animals and Humans*. Cambridge University Press, Cambridge, pp. 61-80.
Gómez, J. C. 1998, “Some Thoughts about the Evolution of LADS, with Special Reference to TOM and SAM”, in: P. Carruthers and J. Boucher, eds., *Language and Thought*. Cambridge University Press, Cambridge, pp. 76-93.
Grice, P. 1957, “Meaning”, in: *Studies in the Way of Words*. Harvard University Press, Cambridge, MA, 1989, pp. 213-223.
Heal, J. 1997, “Radical Interpretation”, in: B. Hale and C. Wright, eds., *A Companion to the Philosophy of Language*. Blackwell, Oxford, pp. 175-196.
Leavens, D. A. and Hopkins, W. D. 1999, “The Whole-Hand Point: The Structure and Function of Pointing from a Comparative Perspective”, *Journal of Comparative Psychology*, 113(4), pp. 417-425.
Morrison, I. 2001, “Attention and Intention”, *Trends in Cognitive Science*, 5(9), p. 375.
Pagin, P. 2001, “Semantic Triangulation”, in: P. Kotatko, P. Pagin and G. Segal, eds., *Interpreting Davidson*. CSLI Publications, Stanford, CA, pp. 199-212.
Picardi, E. 1999, “Sensory Evidence and Shared Interests”, in: M. De Caro, ed., *Interpretations and Causes*. Kluwer, Dordrecht, pp.171-186.
Putnam, H. 1975, “The Meaning of ‘Meaning’”, in: *Mind, Language and Reality*. Philosophical Papers, Vol. 2, Cambridge University Press, Harvard, pp. 215-271.
Samuelson, L. and Smith, L. B. 1999, “Early Noun Vocabularies: Do Ontology, Category Structure, and Syntax Correspond?” *Cognition*, 71, pp. 1-33.
Samuelson, L. and Smith, L. B. 2000, “Attention to Rigid and Deformable Shape in Naming and Non-Naming Tasks”, *Child Development*, 71, pp. 1555-1570.
Talmage, C. J. L. 1997, “Meaning and Triangulation”, *Linguistics and Philosophy*, 20, pp. 139-145.
Tomasello, M. 1998, “Reference: Intending that Others Jointly Attend”, *Pragmatics & Cognition*, 6(1/2), pp. 229-243.

Zaff, B. Z. 1995, “Designing with Affordances in Mind”, in: J. Flach, P. Hancock, J. Caird and K. Vicente, eds., *Global Perspectives on the Ecology of Human-Machine Systems*, Vol. 1, Erlbaum, Hillsdale, NJ, pp. 238-272.

Xu, F. and Carey, S. 1996, “Infants’ Metaphysics; The Case of Numerical Identity”, *Cognitive Psychology*, 30, pp. 111-153.