The Mechanics of the Counterpart Relation

Henk Zeevat*
University of Amsterdam

1 Worlds and minds.

As a rough approximation, we can think of the world as a theory in some logical language, presumably a complete theory. The objects in the world correspond to the individual constants, the properties and relations to the predicate letters etc.

In a equally rough approximation we can think of a mind (or the mind as an information carrier) in much the same way, with the difference that this time the theory need not be complete.

If we take our rough models together in order to get a model of a world and several minds, we need to take care that the non-logical symbols of our languages are distinct. After all, one time a symbol models a bit of John's mind, another time a bit of Tom's, a third time a bit of the world and all these bits are distinct.

The point of these considerations is that the model we obtain is highly incomplete, not in the details of the world or of the minds, but in their connections. What is interesting about minds is that they contain thoughts about the world and about other minds. That means that we must have a relation of reference between minds and the world and between minds among each other. A symbol from Tom's mind may refer to a symbol from the world or from Bill's mind.

This paper is about this relationship, which I will call the counterpart relationship. Notice that this counterpart relation, though employing the same metaphor as Lewis (1968), is nevertheless quite different: we are not relating possible worlds existing independently of each other, but one world and the minds it contains. The situation is like the one in Fauconnier (1985): we have a set of spaces filled with objects and mappings between the spaces. Not all these spaces are mental however: the world itself is an extra one.

The whole purpose of minds is to reflect what the world is like, minds are connected to one particular world by the senses and through action, so that minds are partial causes of events in the world and in other minds. Unlike possible worlds that just exist, the minds and the world are in a constant causal interaction.

It is this causal interaction, which will in the following be the basis of an attempt of explaining how a constituent of some mind can come to be about a part of the world or of another mind.

2 Names

My starting point will be the causal theory of proper names (Kripke 1972). I accept Kripke's arguments against theories which equate the meaning of proper names with a definition, but my

concern in this paper is not the theory of names, but the causality that is assumed in the causal theory of names. The underlying causal principles required for the causal theory of names will be the basis for developing an account of the counterpart relation in the next section.

The causal theory of names can be briefly stated as follows. The fact that a name, say Eliza, in an utterance, say (1),

(1)   Eliza is at home

as uttered now by me to you, refers to Eliza, is caused by some event in the past, the original dubbing of Eliza by that name (or some equivalent phonological string) and the intervening use of the name for that person.

This is a highly reasonable view, but also a mysterious one. It appears to involve a kind of causality whereby a relation between a phonological pattern and an object is caused by the use of language, in particular the use of the phonological form (or its equivalents) as a name for the object.

The claim can be spelled out in the form of the following two causal principles.

(2)   a. if x dubs y N, then N's use by participants in the dubbing will
refer afterwards to y.

b. if x uses N to refer to y in a conversation to z then afterwards
a use of N by z will refer to y as well.

By these laws, we can explain how it comes that Eliza is called *Eliza*, even when everybody has forgotten how Eliza came to have this name. The theory does not rely on a criterion of identification for Eliza, though there may be one or on the ability of users to identify Eliza, though they may be able to do so. Neither does it require that there are users who have the ability to effectively identify Eliza, though again there may be such users. All that the theory gives us is a criterion for the application of the name to its referent in a particular use: Is there a causal link between the current use and the original dubbing?

If we want to say that the name refers to a referent independently of any particular use, it is necessary to assume that the use of the name for the referent has become preponderant in a community of language users. This then allows a certain margin of error that can be corrected by the overwhelming correct use. In principle, however, we can always restrict the reference relation to particular uses of particular users. We will do so in the following.

We will also interpret the theory as giving necessary and sufficient conditions for a name having a referent. If the use of the name is not is not covered by our two principles, it will not refer. This is somewhat unfair, but we are especially interested in seeing where the theory fails to give an explanation.

Let us start with the dubbing principle. The first thing to do is to give a precise determination of a dubbing. Provisionally, a dubbing will be a first use of a name at which the referent is presented: this involves the referent being present and it being clear somehow that the referent is the person intended with the name. Second, mistakes must be discussed. What happens if the name is not new, but already in use? What happens if one of the participants in the dubbing mistakes the presented referent for somebody else ?

It seems that the first kind of mistake does not lead to problems. An old name can be taken up for a new use. There will be a problem in using the name, however, because there will be occasions where the old and the new use are hard to keep apart. The causal theory gives an answer to what happens in the other case: the mistaken participant cannot play the appropriate role in the causal link. But what if the mistaken user just proceeds? She can go on and use the name for

the false referent. Others may start following her. At certain occasions, the false referent may be present. It seems that her mistaken use may like the proper dubbing be a basis for referring to the object and even for the full use of a name in the community. We must, it seems, stretch the concept of a dubbing to any occasion when the name is used in the presence of its apparent referent to take account of this possibility. This will then allow a series of dubbings to be the basis for a reference by the name at a certain occasion. This is not an objection. As long as reference operates correctly, it will be only for theoretical reasons, that we could want to establish that two uses of the name are equivalent in the sense that there is some dubbing that has caused both uses.

The second law needs some loosening up of the concept of conversation. Conversation must include reading, watching television, using the telephone etc. Considerable complexities may be involved here, but I will not go into them. More important is the concept of reference in the formulation. Since we interpret the causal theory as defining reference for names, we have to turn to the theory for a definition of the notion of reference as used in the second law. Dubbing starts the reference to an object by means of a name. The second law carries reference over: exposure to genuine reference allows a person to start using the name for the same referent and to become himself a source for further reference to the same object by the same name.

With respect to the second law, we can also ask what happens when mistakes occur. As it is normally impossible for a language user to determine whether a new name is genuinely referring, it is possible for this user to pick up the name, intending to use it to refer to whatever the speaker was referring to, without the name being genuine: (a) the speaker —knowingly or unknowingly— was using a name without a referent, (b) the speaker may have had a referent in mind without performing a proper dubbing and without relying on an earlier proper dubbing and (c) the speaker may make a mistake: he intends to refer to something, but is actually causally configured to refer to something else by that name. This happens if Bill has acquired the name Mary in the normal way as a name for Mary, but now uses it to refer to Susan because he is under the illusion that "Mary" is Susan's name.

In case (a), the causal theory predicts that the name also does not refer when the new speaker employs it. We were defining what reference is and thereby the conditions for reference are not fulfilled. The question arises whether there are non-referring uses of names, but the causal theory does not say anything about this issue. In case (b), the new speaker will not refer either, whether the speaker had an object in mind or not. Only when the utterance of the speaker is also a dubbing will reference ensue. In the last case (c), we have a more serious problem. Following the causal principle, we will have to say that in this case the new speaker will follow the causal configuration of the speaker rather than the speaker's intention (in the example: the information is about Mary, not about Susan). This is problematic, as there may be clues in the original use concerning the intended reference of the speaker. Having such clues does not turn the use into a dubbing (there need not be clues, the clues need not constitute a presentation of the false referent), but makes it hard to decribe the new speaker appropriately: the new speaker may lean towards the false referent, while being caused to refer to the proper referent. The causal theory does not seem to be able to describe the new speaker's intention of the false referent. The only way out would be to stretch the notion of a dubbing beyond what is reasonable, i.e. by dropping the requirement that the object is presented at a dubbing and replacing it by the requirement that some information is given about the referent. This would make dubbings coincide with any use of the name. How serious this problem is comes out clearly when we realise that for many names, our actual use could well originate in such a mistake, while the original use is lost.

So there are some points where the causal theory fails to predict anything: the cases where the conditions of the law are not fulfilled. Given that names are used in these circumstances and can be used to make true and useful statements, e.g. the well known examples in (3), we have to supplement the causal laws. Also, there are problems in the description of mistakes. The meaning of Santa Claus, Zeus and Vulcan is not explained by the causal theory, since the non-existence of their referents makes it impossible that they have been baptized. Yet all sentences appear to say

something informative.

> (3)　　Santa Claus does not exist.
> Bill worships Zeus.
> John believes Vulcan is covered with craters.

Without disagreeing with the causal theory of names, which does indeed account in an intuitively plausible and essentially correct way for the "standard" use of proper names, i.e. the case where names do refer and originate from a dubbing, we find that the causal principles involved have not been formulated in their full generality. In the next section, I will try to remedy this by stating some principles that govern the construction of the counterpart relation. The causal theory of names will be implied by these principles, but will not itself imply their consequences for the use of proper names.

In the theory we will adopt, the users of a name N associate an internal object with their use of N. The notion of an internal object derives from the classical analysis of representations (see Twardowski (1977)). In any representation, a subject represents something by means of a content. As a representation may or may not represent something outside the mind of the representer, internal objects are necessary. In case the representation is of an outside object, we have a relation between the internal object and the external object. The counterpart relation is precisely this relation. The assumption of internal objects allows us to deal with the cases that are hard to describe if we have only names, referents and users.

3　Some Laws

In this section, I describe the processes that create counterparts in a mind of objects in the world and of the internal objects of other minds.

No doubt the simplest and most common way to acquire an internal counterpart is by seeing objects in the world. If we store our visual experience in memory, we thereby create objects that are counterparts of the objects that affected our senses. We have to distinguish between seeing objects which we recognise (we have seen them before or have other experience of them), and seeing objects which appear to be new. In the first case, the new experience is connected to an old counterpart, in the second case a new counterpart must be created.

When recognition takes place, mistakes can happen. We can attribute our visual experience to a known object that in fact does not cause our experience. An internal object then becomes a counterpart of two different objects: the object seen and the object it was recognised as.

Also when we see a new object, mistakes can occur. We may see an object we are familiar with, but that we do not recognise. In this case, two internal objects will be counterparts of the same external object.
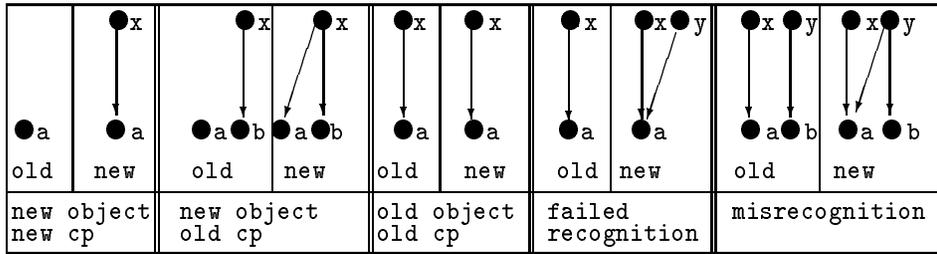
	x		x	x		x	x		x	x	x	y		x	y	x	y
a	a	a	b	a	b	a	a	a	a	x y	a b	a b					
old	new	old	new	old	new	old	new	old	new								
new object new cp		new object old cp		old object old cp		failed recognition		misrecognition									

Fig.1 Changes in the Counterpart Relation

In Figure 1, the five different possibilities are distinguished. Here x and y are mental objects and a and b objects in the world and the arrows indicate the counterpart relation. In each case, the situation before and after the perception is indicated. When a new object is perceived and recognised as new (case 1), in the old situation there is only the object. In the new situation, a new internal object is linked to the object by the counterpart relation. When a familiar object is recognised as such (case 3), nothing changes. The other three cases in the picture are the mistakes: a new object that is recognised as an old one (case 2), an old object that is not recognised (case 4), and an old object that is misrecognised as another old object (case 5).

What holds for visual perception hold for the other senses as well. We can infer a bowl of chicken soup by its smell or by its taste or perhaps by the sound of somebody eating some of it. It would be incorrect, however, to classify counterparts as arising by particular senses, as the most common perceptions are based on combinations of different senses. It is simplest to integrate vision and the other senses and replace the principles for vision by principles referring to a combination of the senses under the label of experience. We obtain the following principles:

(4)   a. An experience of an object that is correctly not recognised by the subject leads to the creation of an internal counterpart of that object.

b. An experience of an object that is incorrectly not recognised by the subject leads to the creation of an extra internal counterpart of that object.

c. An experience of an object that is correctly recognised by the subject leads to the addition of the content of the experience to the old internal counterpart of that object.

d. An experience of a new object that is incorrectly recognised by the subject leads to the addition of the content of the experience to an old internal object which becomes a counterpart of the experienced object.

e. An experience of an old object that is incorrectly recognised by the subject leads to the addition of the content of the experience to another old internal object which becomes another counterpart of the experienced object.

Perhaps some counterparts must be attributed to reason rather than perceptual experience. If there are abstract objects that we can grasp by reason alone, the reasoning process by which we become acquainted with such objects leads to the creation of counterparts for them. This may be the way to account for thoughts about numbers, sets, spaces etc. In that case, however, there

is no ground for giving reason a special status, it comes out as a special kind of experience, i.e. intellectual experience.

Other cases where reason may be responsible for the existence of a counterpart may be illustrated as follows. If I enter my kitchen and see my cat in a typical posture in front of the refrigerator, I immediately infer that there is mouse under or behind the refrigerator. If there is —and my cat has never been wrong in these matters— I may have acquired a counterpart of the mouse by these means, as I have not experienced the mouse myself.

As above, it is reasonable to bring this kind of inference under an extended notion of experience. The only difference is that it is easier to make mistakes in these cases: the postulated mouse has not been seen and need not exist. This is however a question of gradation, since visual illusions, mishearing and other cases of perceptual failure have a similar structure. We can form internal lion-counterparts in response to seeing a wax figure of a lion, we can mistake Paul for Peter, etc. So it does not seem wrong to employ a notion of experience which includes inference from what is directly perceived.

If we want counterparts outside of experience, we can consider innate counterparts. I am not sure it is necessary to assume them, but there seems to be no good reason for ruling them out either. An obvious candidate is the *self*, the privileged counterpart which represents a subject to itself, and which in case of memory loss can get dissociated from the speaker's name. Other candidates are bodyparts and mental experiences like pains.

In the example of the cat, we encounter counterparts of objects in other minds. I infer from the behaviour of my cat that it thinks there is mouse under the refrigerator. The cat's skills in these matters also make me think there is a mouse. If the cat is wrong, I have acquired an internal mouse that is a counterpart of the cat's, without there being an external mouse or another external object of which both the cat and me have a counterpart. Such counterparts arise as soon as minds are involved in the creation of artefacts (here behaviour) from which we can infer that the mind contains certain objects.

If we look at a picture depicting some scene, we get experience which is like normal experience. In some pictorial traditions, the aim in the production of the picture has been to make the experience of the picture as similar as possible to normal experience. As in other experience we form or adapt internal counterparts in response to our experience of pictures. Our counterpart is of the object as represented in the picture, independently of the question whether there is an object that the picture represents. If there is such an external object that the picture represents, our counterparts are thereby also counterparts of the external object that the picture represents. The situation here differs from the case of the cat. We think of the cat's behaviour as a result of the presence of the mouse. The picture however need not be caused by what it represents (though it may). Internal objects resulting from fictional objects are like errors as they are not counterparts of a real object. They differ from errors in that they may do so on purpose. Counterparts of fictional objects may thereby be counterparts of internal objects of the creator of the picture. A principle: an internal object caused by a pictorial representation is a counterpart of whatever the object in the picture represents.

Communication can be compared to pictures. People tell each other about real objects, fictional objects and other objects. Kamp's Discourse Representation Theory (DRT) gives a number of rules for building semantic representations of sequences of sentences (discourses) that define ways of dealing with these reports about objects. The rules in Kamp's theory can be readily understood as analogies for the rules for perception that we have considered until now. Their aim is different however as Kamp tries to make sense of the linguistic distinctions in referring expressions (definite versus indefinite, perfect versus past tense etc.) Because of this aim, the representations represent not so much what actually happens in a hearer, but the development of the hearer as expected by the the speaker. Here we are interested in what really happens to the hearer. As a result we will have to modify the rules.

The following two rules are adapted from Kamp to our current context. The rule for definites is a very rough approximation.

(5)　　a. If an indefinite expression is encountered, create a new internal object as the reference of the indefinite that is a counterpart of the internal object of the speaker that is the speaker's reference for the indefinite.

b. If a definite expression is encountered, identify an old internal object meeting the description as the object to which the definite refers as a counterpart of the speaker's reference.

The last clause covers names, but names cannot always be identified with old objects of the interpreter. When the name is used for the first time the object to which the name refers must be created. As is well known, for many definites the same rule obtains: if they are not identifiable in the representation of the context, a counterpart must be created. We will have more to say about this issue later on.

Similarly, we must also take into account the situation where an indefinite is used (indicating that the speaker assumes that the hearer will not already have an internal object for it) but that the hearer nonetheless recognises the object to which the speaker refers. In that case an old counterpart is used.

This means that we can state the principle in a rather simple way:

(6)　　a. If an expression is encountered that refers to an object and the hearer does not know which object this is, the hearer creates a counterpart of the speaker's object that caused the speaker to use this referring expression.

b. If a referring expression is encountered which is recognised as referring to whatever an internal object of the hearer refers to, the hearer's internal object becomes a counterpart of the internal object of the speaker that caused the referring expression.

Definiteness and indefiniteness are not decisive in this context, but the distinction can still be interpreted as an indication of the strategy that the speaker thinks the hearer must follow in dealing with a referring expression.

Communication builds counterparts to the internal objects of the speaker, internal objects, which in turn may be counterparts of an external object. This means that we must consider not only referential mistakes, but also the mistakes that a subject has inherited from other speakers by communication.

The following two examples illustrate ideal communication, i.e. communication according to the speaker's assumptions.
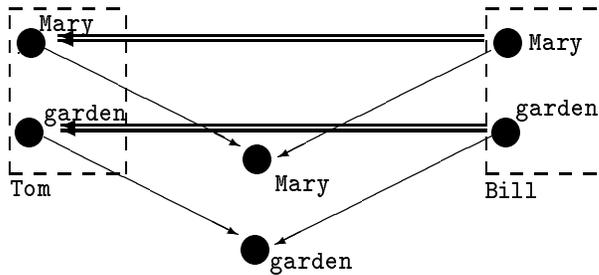
Fig. 2 Tom says: Mary is in the garden

Figure 2 presents an ideal successful utterance of the sentence: Mary is in the garden. Bill has a counterpart of Mary and of the garden and identifies them with the objects Tom is referring to. The bold arrows indicate the possibly new elements of the counterpart relation: Bill's object becomes linked to Tom's.
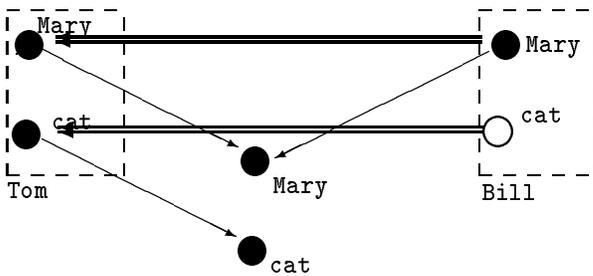

Fig. 3 Tom says: Mary saw a cat

Figure 3 presents an equally ideal utterance of *Mary saw a cat* to Bill who is not yet acquainted with the cat. So Bill's counterpart of the cat is new and Bill is only indirectly related to the cat.

When a new name is used in communication, the hearer acquires more than just a counterpart. With the name, he also receives a means for referring to the object he has acquired a counterpart of. For him and for minimally the speaker, the name will be a means for getting back to the object in question. The name is associated with an internal object, by the communication process the internal object is a counterpart of an internal object of the original speaker and thereby of whatever that object is a counterpart of.

This process takes the place of the causal theory of reference. The acquisition of a new (use of) a name is the association of the name with an internal object which is a counterpart of other objects. In a dubbing, the object is presented and by its presentation participants acquire a counterpart of the object. The name thereby becomes a name for the presented object for those involved in the dubbing, at least if the presentation is successful.

The second law is also a direct consequence of the principles sketched here. The counterpart relation is established between the new user's internal object and the speaker's internal object. If the speaker's object is a counterpart of an external object (through other use and a dubbing) the new user's object will be a counterpart of that same object.

In contrast to earlier on, there is now a fairly clear account of what goes on in the marginal cases. The new user always acquires a counterpart of the speaker's object and is thereby is able to discuss that object with the speaker, independently of any external object present. In the case of mistakes, the counterpart relation will split in different ways.

The explanation of language use involving names without a denotation is unproblematic and we

can even explain how such names can be used to make true statements. Consider e.g. (7).

(7)   Santa Claus does not exist

We can form a counterpart of Santa Claus by any of talk of Santa Claus. We can then deny that there is a real object which underlies this talk. Or consider the following example, involving the non-existent planet Vulcan.

(8)   Harry believes Vulcan is covered with craters

We can pick up a counterpart of Vulcan by the talk of this planet. Harry's Vulcan must be related to the same talk, and the sentence reports a belief of Harry about this object.

4   Some Concepts

The picture so far suggests some alternative conceptions around traditional concepts like thoughts and the world. In this section I will develop these to a minimal extent. The idea is to have a large class of entities of the type usually called proposition, thought, or fact etc. I will call them *thoughts*, though potential meanings of sentences is probably a more neutral description. Thoughts will include mental representations, but they need not be mental representations: typically they are whatever a mental representation can be a counterpart of. Within the thoughts, we single out some special classes, facts, possible facts, private thoughts, proper thoughts, etc. The construction of the class of thoughts involves one of the many ways now available for building structured propositions. Any method will do, as long as it allows us to recover the basic building blocks of the structured object and its logical properties.

First of all, the notion of counterpart must be extended to kinds and universals, (in principle to anything which is in the world). In experience of kinds and universals, counterparts of real kinds and universals are built up, in communication involving them further counterparts come into being. Again, innate properties may be assumed.

Second, we assume some theory of structured propositions. Now, universals and internal universals may be combined with objects and internal objects. In this way, we can form propositional entities out of the various entities we have so far considered.

This allows the following definition of a thought.

A *thought* is a propositional construction out of internal and external objects and internal and external universals.

Within the thoughts, we can consider *external thoughts*, the ones completely constructed from external objects and universals.

External thoughts are the possible facts: they can be in the world or not.

Facts are the external thoughts that are in the world.

Private thoughts are thoughts made up entirely from the internal objects and internal properties and relations of one particular mind. A mind can be equated with the collection of the private thoughts of that mind. Mental representations are private thoughts which form the stuff that minds are made from.

Most interesting are, however, thoughts made from different sources at the same time, the so-called *mixed thoughts*. These are the proper thoughts: They are objective without necessarily being possible facts. They are subjective without coinciding with representations.

The counterpart relationship can be generalised to thoughts:

> (9)   a. One thought is a counterpart of another thought, if it can be
> derived from the other thought by the substitution of a constituent
> object by its counterpart.
> b. A thought A is a counterpart of a thought C if there is a thought
> B such that A is counterpart of B and B is a counterpart of C.

In terms of the counterpart relationship between thoughts we can state what it means for two minds to agree with respect to some thought: both minds have a counterpart of that thought.

Also, we can define beliefs:

> (10)   A thought is a *belief* of a mind iff the mind has a counterpart of
> the thought.

And in the same way, truth for thoughts.

> (11)   a thought is *true* iff it is a counterpart of a fact.

It is however not clear that this covers all relevant situations. Suppose I guess that there is person waiting in the hall, without any direct evidence for this, just based on my general expectations. Suppose furthermore that there is a person waiting. It then holds that my thought that there is a person waiting is true, even if my thought is not a counterpart of the actual situation.

There are two solutions for this problem. The first starts by noting that it is not necessary to think of the thought as containing a counterpart of the waiting person himself. The thought may be built from other objects, like the universal *person* and the universal *waiting*. The private objects in the mind are then true counterparts of these universals. (This would be in the style of Russell's treatment of definites.) There is no reference to the person, but there is a genuine quantification going on. Definition (11) will continue to hold.

The second route is to allow extensions of the counterpart relation. Where a private object — the waiting person in my thought in the example— is not a counterpart of anything at all, the counterpart relation can be reasonably extended to include a link between the private object formed in a mind and a real object. In this way, we can say that my thought is true because of a reasonable extension to the counterpart relation. This is in the spirit of Kamp's definition of a truthful embedding in Kamp (1981), which can be taken to consider all ways of establishing a counterpart relation between abstract representations and the world.

Note that one can still debate cases like the following. Suppose, I think that John is waiting in the hall. On the basis of this thought, I also think that there is a person in the hall. Suppose that there is a person in the hall, but Bill, not John. My first thought is false. But what about the second, derived, thought? My intuition here is that unless I would accept the thought that John or somebody else is waiting, the thought is still false. So perhaps extensions must be free: we can link an internal object to an external object only if the internal object is not linked yet.

Extensions can also be considered for a definition of agreement between people and for belief in a thought.

5   Interpreting Belief Ascriptions

When do we think that a belief report is true? We hear an utterance, e.g. *John believes that Fido is on the mat*, and interpret the complement of that sentence. This interpretation constructs a

private thought of the interpreter based on the use of the names of both objects and universals in the sentence and the forms we associate with the remaining material. There is nothing special about this interpretation: the fact that it is a complement of a belief sentence does not make much difference. The one exception seems to be the treatment of possible presuppositions of the complement: they are evaluated against the context of the beliefs of the belief subject which may contain material not contained in the basic context.

The private objects and universals in the thought so formed are for all the interpreter knows counterparts of real objects: this is a precondition on the successful use of names. (We shall see later on that this principle can be specialised to other contexts.) Thereby the private thought of the interpreter appears to him as a counterpart of some other thought. It is of this other thought that the interpreter is informed that John's mind has a counterpart. The interpretation process is thereby a vital intermediate step in determining the content of a belief ascription.

The theory we have established so far, without embarking on the subtleties of interpretation, is already able to analyse certain puzzles in the area of belief sentences. This is not the place to embark on a lengthy discussion of difficult cases in the theory of belief sentences, but they may serve as an illustration of the general concepts I introduced. Figure 4 illustrates an approach to Kripke's puzzle (Kripke 1979), taken here as the problem of explaining the possibility of having a consistent set of beliefs that includes two beliefs that taken as propositions are contradictory.
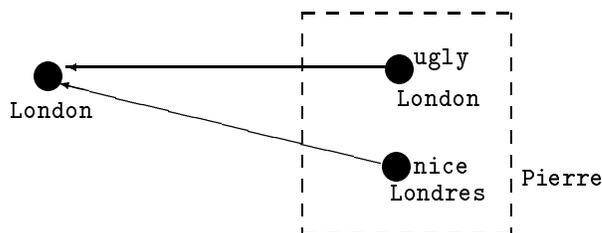

Fig.4 Pierre and London

The propositions "Londres is nice", "London is not nice" are contradictory, as the names *London* and *Londres* refer to the same object in the world. For Pierre, however, there are two counterparts of this object, associated with *Londres* and *London* respectively. So his beliefs are consistent in this respect.

The second example (figure 5) illustrates the Hob-Nob example of Geach (Geach 1967).

(12)   Hob believes that there is a witch in the village and Nob believes
         that she poisoned his cow.

In scope ambiguity treaments, only *de re* readings can be admitted here. This is contrary to our intuitions, as we do not feel that in order to understand (12) we need to believe in the existence of witches. Here, we have two ways of constructing a reading. The first one is by taking the witch to be invented by Hob, who has communicated with Nob, thereby causing Nob to form a counterpart by communication. Intuitively, it is not required that Hob has talked to Nob, it is sufficient that both have been exposed to the same rumour, originating with some person x. This is illustrated in figure 5.
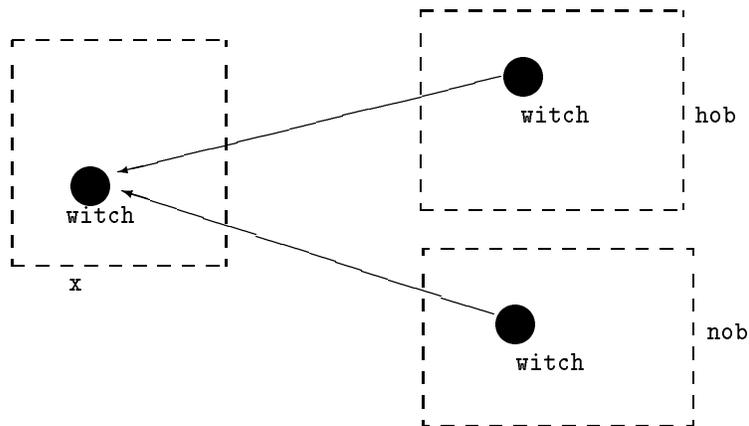
11

Fig. 5 Hob, Nob and the Witch

There are however problems that are not so easily dealt with. The classical one is known as Frege's puzzle, as it derives from Frege (1892).

    (13)   The Babylonians believed that Hesperus was different from Phos-
           phorus.

(13) is true, but (14) is false.

    (14)   The Babylonians believed that Hesperus was different from Hespe-
           rus.

On the assumptions we have made so far, it should be true that (14) is true if (13) is true. The Phosporus counterpart of the Babylonians is a counterpart of Venus and can thus be named Hesperus by us even if the Babylonians would not be prepared to do so.

That this last assumption is correct can be seen by another example. We can without any problem report that John thinks that Bill is wounded if John —who has never seen or heard of Bill— finds Bill lying next to his bicycle. It cannot be a question of the belief subject knowing the name, whether or not we can use the name in an ascription.

The remainder of the paper will be devoted to a solution for this problem. This solution is rather linguistic and makes use of recent studies of presupposition, i.c. the anaphoric theory of presupposition, developed by Van der Sandt (1989). My presentation of this theory is based on Zeevat (1992) and imports some ideas of Heim (1983) to the original theory.

5.1   Presupposition Theory

The material to which the speaker of a sentence is committed can be divided in two parts: a part which is assumed at the outset of the utterance and a part that is added by the sentence as new information. There are a number of devices that make material presupposed: lexical items, syntactic constructions and intonation. These devices are called presupposition triggers. For the interpretation process, it makes a considerable difference whether material is presupposed or asserted. Presupposed material can either be identified in the context or be accommodated at appropriate places in the context. Asserted material can never be identified and is always added to the context to which it belongs syntactically.

To make this precise, some notion of context and of a hierarchy of contexts is required. New contexts are standardly introduced by negation, implication, quantification and attitudes. In this way, we can distinguish the following six contexts in the sentence (15).

(15)   If John believes that Harry is a bachelor, he does not believe that
Bill is.

Contexts are generated by logical operators and have a content which can be read of from the operator. The following table lists the six contexts of (15) with the responsible operator.

C_0		the basic context
C_{00}	if	the condition of the conditional
C_{000}	believe	belief complement
C_{01}	then	the consequens of the conditional
C_{010}	not	scope of negation
C_{0100}	believe	scope of second belief

The content of the contexts is determined by what is already known. In C_0, we find what has been established in the preceding conversation, in C_{00} we assume the content of C_0, in C_{01} the content of C_0 plus the content of the if-clause, in C_{000} what is believed by John in C_{00}, in C_{010} (most of) the content of C_{01} and in C_{0100} what is believed by John in C_{010}. Negation (like the condition of a counterfactual) has the special property that it inherits the content of the embedding context only in so far as this allows a consistent construal of the new material to be inserted there. This property is necessary for an adequate treatment of negations involving non-existent objects.

The content of the context determines whether the presupposition can be found or can be added. It can be found if it is present in the content of the context or in the content of superior context, it can be added to a context if it is consistent with the content of the context. A version of Van der Sandt's theory lets the process search the content of the presupposition from the local to the global context. The local context is the syntactic context of the trigger (the smallest scope of an operator in which the trigger occurrence takes place). If it is necessary to extend the search to a higher context, the presupposition is added to the context that was searched, unless the addition would cause an inconsistency there. In case we cannot find the presupposition, the process breaks off at the highest level with another default addition of the presupposition. Finding the presupposition is called resolution, adding it accommodation. If both resolution and accommodation consistently fail for all contexts, interpretation is blocked.

Proper names trigger two presuppositions: an existence presupposition and the presupposition that the object carries the name. The existence presupposition is the claim that there is something to which the name refers. The name binds an argument of some predicate. The existence presupposition is the claim that that argument exists[1]. The second presupposition is the statement that the argument is the referent of the name. In a DRT-like formalism the first presupposition is the statement that x is a discourse marker, the second the statement $x = n$. Classical approximations are given in (16),

(16)   a. $\exists x\, x = y$
         b. $y = n$

---

[1] The position, which some have defended, that the existence presupposition is triggered by the predicate rather than by the name, is by itself quite reasonable. It runs into problems however when we are dealing with argument positions that allow non-existing objects, e.g. :

        Frege discussed Vulcan.

If we replace *Vulcan* by an unfamiliar name, say *Eliza*, then, in the default interpretation, Eliza exists. Making the predicate the trigger for the existence presupposition is impossible in this case. The default interpretation can only be explained by assuming an existence presupposition on names.

where y is a free variable for the position that the name occupies.

We have already seen why we need a double presupposition. If Harry believes Bill is wounded, without knowing that Bill is the name of the unfortunate victim, Bill's beliefs include the existence presupposition, but not the name-presupposition. In making the report, we assume both to be part of the global context. The two presuppositions do not have the same status. The name-presupposition makes it possible to identify the referent among others, the existence presupposition guarantees that it refers in its context. Thereby, it is a necessary condition on the local context that the existence presupposition holds, but not that the name-presupposition holds. At the same time, the name presupposition entails the existence presupposition: you can have a name only if you exist. A name without a referent is a name that is limited to a subordinate context.

Let us first apply the above to a simple example.
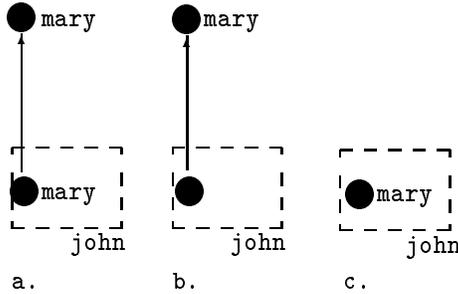
(17)   John thinks that Mary sleeps.

Fig. 6 John thinks that Mary sleeps

The interpretation process lets us to deal with two presuppositions and two hierarchically ordered contexts: C_0 and C_{00}, where the second contains what John believes according to C_0.

The result of the presupposition treatment will be:

(18)   a. if C_0 contains a Mary or allows a Mary and C_{00} allows a Mary
the result is:
both C_0 and C_{00} contain a Mary where C_{00}'s Mary is a counterpart of C_0's one.
b. if C_0 contains a Mary or allows a Mary and C_{00} does not allow one, the result is:
C_{00} contains a counterpart of C_0's Mary (which is not a Mary).
c. if C_{00} contains a Mary, C_0 disallows one, the result is:
C_{00} contains a Mary which cannot be a counterpart of a Mary in C_0 (but possibly of something else)
d. if neither C_0 nor C_{00} allow a Mary, interpretation fails.

Of course it is possible to give approximations of the propositional content. For (d) there will be no content, for (c) we have to split the content in (c.1.), where an external object (not called Mary) is assumed and (c.0.) where no such counterpart is present. (c.0.) is a counterpart of the classical *de dicto* reading, (b) of the *de re* reading.

(19)   a. $\exists x(x = mary \land believe(j, \exists y(y = x \land y = mary \land sleep(x))))$
b. $\exists x(x = mary \land believe(j, \exists y(y = x \land sleep(x))))$
c.0. $believe(j, \exists y(y = mary \land sleep(y)))$
c.1. $believe(j, \exists y(y = x \land y = mary \land sleep(x)))$

14

The innovative readings are (a), the default reading where we assume internal and external Mary's where the first is a counterpart of the second and (c.1.) which is one of a range of intermediate readings.

5.2 The problem resolved

Consider what happens in the Babylonian example. In (13) we have the task of both resolving Hesperus and Phosphorus from the belief context. Phosphorus (unlike *itself* or Hesperus) cannot be locally resolved to the Hesperus counterpart of the Babylonians, as *Hesperus* does not meet the name-presupposition of *Phosphorus*. So by default we have internal and external Hesperus's and Phosphorus's linked by the counterpart relation, where the external objects coincide and the internal ones are different. This gives us consistent beliefs about the same object, that externally would be inconsistent.

In (14), the second occurrence of Hesperus will be resolved to the object to which the first occurrence refers: the names are identical. This leads to an inconsistent context, as the context denies their identity. As this is resolution and not accommodation, the process does not need to preserve the consistency of the context. This interpretation explains the intuition that we are here dealing with the attribution of an inconsistent belief to the Babylonians.

The same explanation will account for an asymmetric version of the problem, as in (20).

> (20)   a. The Babylonians believed that Phosphorus was visible in the morning sky.
> b. The Babylonians believed that Hesperus was visible in the morning sky.

In (20a) , a truth is expressed. (20b) should be true by the reasoning that we followed before: Hesperus is the name of an object of which the Babylonians believe that it is visible in the morning sky. However, by default we obtain the name-presupposition in the belief-context, which leads to an interpretation that is false as we cannot find an object in the Babylonian mind which is both visible in the morning sky and Hesperus. The only way to overcome this interpretation is by assuming that the name-presupposition is false in the context of the Babylonian beliefs. In that case, we indeed get an interpretation under which (20b) is true, again in accordance with our intuition that holds that (20b) is true in one sense and false in another.

The explanation uses a distinction between the apparent logical form and the real logical form defined by a process of interpretation in which presupposition are resolved or accommodated, according to the content of the context of their triggers. The sentence, understood as a formula with a structural similarity to the sentence, determines one truth condition. The interpretation process may however assign a different structure to the formula that characterises the interpreted sentence.

The resolution and accommodation processes lead to extra constraints on the counterpart relation: resolution by unifying the internal object with an internal or external one so that the internal object must be a counterpart of or identical to the object to which it is resolved (resolution within the same world or the same mind means identity), accommodation by creating an external object of which the internal object is a counterpart.

The above gives an account of the Fregean puzzle without assuming senses. We can be pure Millians and still explain the semantic puzzle: why is there no *substitution salva veritate*. (substitution does not preserve the interpretation of the sentence). Also, we have a basis for an attack on the epistemic version of the puzzle: we describe the Babylonian mind in such a way that it will change under the influx of the new information that Hesperus is Phosphorus. We do not need

senses, as we can describe what change a statement can make to a mind and that is all we need for explaining Frege's intuition that identity statements can be informative.

A Fregean may retort that, now, we can associate with the linguistic entities partial functions mapping minds (or information states) to new minds. If necessary, we could consider this function to be the sense of the expression and define the meaning of complex expressions by function composition. Nobody can —it seems— quarrel with this alternative account of senses. Fregeanism, however, goes beyond this simple view of senses by postulating that sometimes the senses themselves are the referents. It is here that we are forced to disagree.

5.3   Definites and Indefinites

The Millian account can be extended to cover other referring expressions as well. If we believe -with Russell- that for definite descriptions there is no Fregean puzzle, because the content of the description belongs to the assertion that is made, we obtain a model to which we can then try to assimilate names and demonstratives. Much work has been devoted to such attempts and to showing that such solutions cannot work.

What I want to propose is the reverse: a Millian account of definites and indefinites. I believe such an account is important for a better understanding of what goes on in belief contexts and for a better grasp on incorrect, but successful reference by means of these descriptions. There is an important proviso here: I do not want to claim that all uses of descriptions must be so understood. In particular, there are clear uses of both definites and indefinites where they contribute to the assertion. For example, both types of descriptions can be used in predicative positions and they can both be involved in quantificational structures. Nonetheless, an important use of these descriptions is to refer to objects. A direct reference account of definite descriptions has been provided in Kaplan (1978), here we extend it to indefinites.

Reference to objects is a special kind of action. As other actions, it is only successful if certain action presuppositions are fulfilled. In this case, an obvious presupposition is that there is something to which the expression can refer. The second action presupposition is that the means chosen for referring are adequate. The rules involved here are complex, but often have to do with properties of the context that could not possibly contribute to the content of the utterance (e.g. the referent has been mentioned before, it is salient in the context, etc). For descriptions, one condition is always that the referent meets the description. Bearing the name is the analogous condition for proper names.

What must meet the two presuppositions is the context of the trigger (here the referring expression). In simple belief sentences, typically, a trigger under the belief will have its presuppositions satisfied in both contexts: the basic context and the context given by the belief operator. Resolution and the presence of incompatibilities will result in weaker readings.

What happens if a reference is made, but one of the presuppositions fails to hold ? For existence, it will lead to misinformation: the interlocutor will assume the existence of the referent and consequently obtain information that the speaker cannot guarantee. For adequacy, there is a danger of false information arising or the identification of a wrong referent. Similar effects arise if the conditions of use are transgressed. In all cases, identifications that should have been made are not made, unintended identifications may take place and false information may be accommodated. All in all, the speaker does not get her information across.

This leads to the assumption, that the speaker, in choosing his means of reference, indicates with respect to each of the references she is making that she thinks the object exists, that the object meets a certain predicate or bears a certain name, etc. This information is over and above what the speaker asserts in her utterance or at least may be so: of course, the speaker knows that she is conveying extra information and may exploit this fact.

In the local contexts arising from quantification, negation and the attitudes, antecedents for the presuppositions may be in the local context or in an intermediate context or in the outermost context. If the antecedent is not found in the local context, the identification of the antecedent is accompanied by accommodations in intervening contexts. When there is no antecedent, accommodation takes place in all contexts of the trigger. Accommodation is here attempted addition: addition takes place when it does not yield an inconsistent context.

In figure 7, a standard definite description gets three readings: a double accommodation (accommodation in the outside box may be resolution) (a), a *de re* reading (b) (John knows his supervisor but not as his supervisor) and a *de dicto* reading (c) (John has no supervisor but only thinks so).
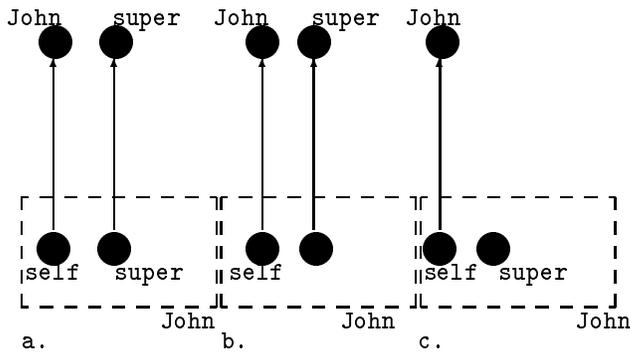

Fig.   7 John thinks that his supervisor hates him

At this level of description, completely the same happens with indefinites as in figure 8.
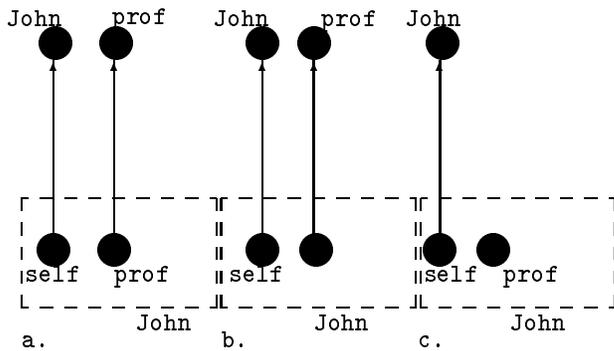

Fig.   8 John thinks that a professor hates him

Here we are dealing with a case which generally is not considered to be a presupposition, but which nonetheless behaves much like one. An obvious reason for not treating the existence presupposition and the descriptive one we associated with indefinites as presuppositions is that their presupposition is not supposed by the speaker to be known to the user. The use of an indefinite indicates that the presuppositions cannot be resolved to the context. But the presupposition is projected out of large number of contexts, including quantification and negation. We all know the examples: they are the classical examples of quantifier scope ambiguity. Compare (21):

(21)   *John does not know a girl*

Here the most plausible interpretation is the one obtained by accommodation (the other interpre-

tation is better expressed using *any*). What I have to explain is why this is not always the default reading. But before discussing that, we can observe that the presuppositions of indefinites may break out of any operator containing them.

To explain why they do not always break out of the scope of a quantifier, it is necessary to introduce a general property of presuppositions. We have seen two cases where projection is blocked: resolution and incompatibility with the exterior context. There is a third factor: when they contain a variable that is bound by the quantifier.

> (22)　　a. Every boy regrets that he lost his bike.
> 　　　　b. Whenever John buys a car, there is something wrong with the
> 　　　　motor.
> 　　　　c. When John drinks, he has a beer.

(22a) illustrates that a definite (his bike) can be trapped inside the scope of a quantifier and so does (22b) . The explanation here is that there is a dependence on the boy (the car) in the condition. (22c) illustrates the same for indefinites. The explanation here should be the same: a dependence on the occasions of drinking that the conditional quantifies over.

(23)illustrates some other positions that indefinites find it hard to escape from.

> (23)　　a. Every boy who knows a girl loves her.
> 　　　　b. Every boy who knows the girl loves her.
> 　　　　c. When a tourist visits a museum, he is impressed.
> 　　　　d. When a tourist visits the museum, he is impressed.

There is a strong contrast between (23ac) and (23bd) . Though further study of this phenomenon is certainly necessary, we may speculate that indefinites have a tendency to become dependent on the quantified variables active in their context[2]. As in the case of dependent definites, this limits their accommodation to the scope of any quantifier on which they have become dependent. As beliefs are not quantifiers, this would explain why indefinites tend to float out from belief contexts.

Another argument for the view that indefinites are (accommodation only) presuppositions comes from languages such as Latin or Russian which do not mark the grammatical distinction between definites and indefinites.

> (24)　　Homo currit. (*man runs*)

(24) can mean that a man runs, but also that the man runs. We should describe *homo* as an NP as a presupposition trigger with almost standard behaviour: it tries to be resolved and if it fails it gets accommodated. This allows a description of the bare singular NP in Latin which just says that it presupposes the existence of something meeting the content of the noun. This contrasts with the specialisation we find in our languages. Short deaccented definite descriptions must be resolved (directly or by bridging), long ones can also identify referents in the non-linguistic background (possibly by bridging or accommodation) and indefinites, finally, must be accommodated. Each of these specifications is a separate rule.

---

[2]The phenomenon of *er*-insertion in Dutch with indefinite non-subsecting NPs points in the same direction: the referent of the indefinite needs to be located. If we assume that the special contexts all introduce a location, we have our explanation.

6 Some considerations

This is of course a rather reactionary paper which goes against the successful Fregean revolution in semantics in holding that reference is primarily a property of minds: the successful intention of external objects. Traditionally, language is only a means to transfer representations from one person to another. Our only addition to pre-fregean semantics is the observation that this process will also transfer reference.

Where do we disagree with Frege? In the first place, in the theory of reference. Here, the only names without a reference are the ones where the name is only a *flatus vocis*, i.e. even the speaker does not intend anything with it. As soon as the speaker has an internal object for it, it can be used as a name. It is in no way guaranteed however that a true or false thought can be communicated by means of it. Much the same holds for words that refer to universals.

Also, there is no obvious motivation for a theory of sense. We have seen how Frege's problem can be solved without any appeal to a theory of sense. We only required the assumption of an interpretation mechanism. This motivation is also not forthcoming from studying indirect contexts, as the natural way of interpreting them will be their interpretation in other contexts (e.g. what we know about another person's beliefs) than the current one.

A person who is able to incorporate the information presented by some linguistic utterance into her existing information (or the information she assumes others to have) in accordance with the use of the expression involved in the utterance, has sufficient grasp of the meaning for the task at hand. If the person is also capable of correctly formulating information at her disposal by the expression, we have all that is required for correct use of the expression. Though we may want to reconstruct meaning as a function from information states to information states, it does not follow that this notion has any other role then characterising how expressions are used, i.e. provide a criterion for when a person uses an expression correctly and when not. The knowledge of the function as such goes beyond what is required by communication, since all we require of the subject is how to knowledge with respect to the information states she has at her disposal.

References

Fauconnier, G. (1985) *Mental Spaces: Aspects of Meaning Construction in Natural Language.* MIT Press, Cambridge (Ma.).

Frege, G. (1892) Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* **100**, p. 25-50.

Geach, P. T. ((1967) Intentional Identity. In: *Journal of Philosophy* **LXIV** p. 627-632.

Heim, I. (1983) *On the projection problem for presuppositions*, *WCCFL* **2** p. 114-26.

Kripke, S. (1972) Naming and Necessity. In: D. Davidson & G. Harman (eds.) *Semantics of Natural Language*, Reidel Dordrecht p. 253-355.

Kripke, S. (1979) A Puzzle about Belief. In: A. Margalit (ed.), *Meaning and Use*, Reidel Dordrecht p. 239-283.

Lewis, D. (1968) Counterpart Theory and Quantified Modal Logic. In: *The Journal of Philosophy* **65** , p. 113-126.

Kamp, H. (1981) A Theory of Truth and Semantic Representation. In: J.A.G. Groenendijk, T.M.V. Janssen & M.B.J. Stokhof (eds.). *Formal Methods in the Study of Language.* Mathematical Centre, Amsterdam p. 277-322.

Kaplan, D. (1978) Dthat. In: P. Cole (ed.) *Syntax and Semantics 9: Pragmatics.* New York, Academic Press, p. 221-243.

Van der Sandt, R. (1989) Presupposition and discourse structure. In: Bartsch, van Benthem, van Emde Boas (eds.)*Semantics and Contextual Expression.* Foris, Dordrecht.

Twardowski, K. (1977) *On the Content and Object of Presentations.* Nijhoff, The Hague.

Zeevat, H. (1987) A Treatment of Belief Sentences in Discourse Representation Theory. In D. de Jong, J. Groenendijk & M. Stokhof (eds.) Studies in Discourse Representation Theory and the Theory of Generalised Quantifiers. Foris, Dordrecht, p. 189-215.

Zeevat, H. (1992) Presupposition and Accommodation in Update Semantics. In: *Journal of Semantics*, **9**, p. 379-412.