Advances in Cognitive Systems: Papers from the 2011 AAAI Fall Symposium (FS-11-01)

Protocols for Reference Sharing in a Belief Ascription Model of Communication

Yorick Wilks
Florida Institute of Human and Machine Cognition
15 SE Osceola, Ocala FL 37771
ywilks@ihmc.us

Abstract

The ViewGen model of belief ascription assumes that each agent involved in a conversation has a belief space which includes models of what other parties to the conversation believe. The distinctive notion is that a basic procedure, called belief ascription, allows belief spaces to be amalgamated so as to model the updating and augmentation of belief environments. In this paper we extend the ViewGen model to a more general account of reference phenomena, in particular by the notion of a reachable ascription set (RAS) that links intensional objects across belief environments so as to locate the most heuristically plausible referent at a given point in a conversation. The key notion is the location and attachment of entities that may be under different descriptions, the consequent updating of the system's beliefs about other agents by default, and the role in that process of a speaker's and hearer's protocols that ensure that the choice is the appropriate one. An important characteristic of this model is that each communicator considers nothing beyond his own belief space. A conclusion we shall draw is that traditional binary distinctions in this area (*like de dicto/de re* and attributive/referential) neither classify the examples effectively nor do they assist in locating referents, whereas the single procedure we suggest does both. We also suggest ways in which this analysis can also illuminate other traditional distinctions such as referential and attributive use. The description here is not on an implemented system with results but a theoretical tool to be implemented within an established dialogue platform (such as Wilks et al. 2011).

Introduction

Since an addressee's beliefs are not directly available to a speaker nor *vice versa*, the problem for any computational reference theory applicable to human communication is how a question is to assert that there is an external object denoted for both the speaker and the addressee and how, by choosing a referring expression which denotes this external object, the speaker can cause the addressee to pick out the object he intends. Traditional approaches include (Kronfeld 1990) and (Perrault & Cohen, 1981). Such theories traditionally have difficulty with reference to non-existent or dubious objects such as Santa Claus, UFOs, or the vision I had yesterday, but years of ingenuity have produced accounts of these, however tortured, within a realist framework. However, far greater difficulties are presented by situations where the speaker and hearer simply do not agree about what entities there are out there, such as when one ship's captain believes two naval entities are the same and another other does not, yet they have to communicate. We gave an account within the ViewGen belief ascription framework some time ago: (see Ballim & Wilks, 1991) of how effective discourse entities could be constructed in such cases, as well as how that account could be procedurally unified with an account of metaphor, an apparently distinct phenomenon.

Following the classic examples in this field, one could say that, when Smith asserts to the police:

My wife's murderer came in by the back door and tied me up first

those hearing the sentence may perform wide range of referential procedures. Some may construct an explicit intensional entity corresponding to the description in the first three words and ascribe to it whatever other beliefs about suspects they currently hold. Others, already convinced of Smith's own guilt, and hence the literal untruth of that statement above, may construct conflated intensional entities of the sort we have described elsewhere (Wilks, Barnden and Wang, 1991) and which we could call *Smith-as-wife's-murderer* and then attempt to ascribe (cynical and devious) beliefs to that intensional entity based on beliefs they already hold about the components of the hyphenated composite. In neither case do traditional distinctions like *de dicto/de re* have any strong role to play, a point we shall make in more detail later in the paper. The form of this paper is not a report on work accomplished but an investigation of reference within the ViewGen paradigm, together with a proposal to install it within an established dialogue management platform, such as the author's Companions platform (Wilks et al., 2011).

Our mode of description here is that of agents participating in real-time dialogue, but the issues involved in text understanding are no different, and to some are more pressing. In an example like the following:

A Spanish priest was charged here today with attempting to murder the Pope. Juan Fernandez Krohn, aged 32, was arrested after a man armed with a bayonet approached the Pope while he was saying prayers at Fatima on Wednesday night. According to the police, Fernandez told the investigators today he trained for the past six months for the assault. He was alleged to have claimed the Pope `looked furious' on hearing the priest's criticism of his handling of the church's affairs. If found guilty, the Spaniard faces a prison sentence of 15-20 years.

(an example due with thanks to Sergei Nirenburg from the London *Times*)

The six underlined phrases all refer to the same man, a vital fact for a computer understander or translator to know, since some of those phrases could not be used in any literal manner in another language (e.g. Spanish). Our claim is that references like these can only be related appropriately by some heuristic system like the one we describe here, based on the most plausible matches within belief spaces, in this case of a reporting journalist who is probably not a Spaniard, for example.

The basic idea behind the ViewGen model is that each agent involved in a conversation has a belief space which includes models of what other parties to a conversation believe. The distinctive notion is that a basic procedure called belief ascription (Wilks & Ballim, 1987; Ballim & Wilks, 1991) allows belief spaces to be amalgamated on demand so as to model the updating and augmentation of belief environments. In this paper we extend the ViewGen model to a more general account of reference phenomena, by using the notion of an reachable ascription set (RAS) that links intensional objects across belief environments so as to locate the most heuristically plausible referent at a given point in a conversation.

The key notion is the location and attachment of entities under different descriptions; the consequent updating of the system's beliefs about other agents by default, and the role in that process of speaker's and hearer's protocols that ensure that the choice is the appropriate one. These protocols are intended to make an agent's own beliefs more accessible to another (on the assumption no deception is involved). The important characteristic of this model is that each communicator considers nothing beyond his own belief space. Whether what he believes the other believes is in fact what the other party actually believes, or whether the utterance is true or false relative to the external world, is not directly accessible to him.

Background

Although this work has been a developing paradigm for some years, it does have obvious affinities with other AI traditions, particularly the early work of Maida (1984) on ordering referential candidates within belief environments. Clear differences with Maida are that he postulateed single denotational entities for certain classes of intensional objects, that are equivalent in some belief space, which has the, for us, unacceptable consequence of requiring real-world equivalence outside belief spaces. His assumption removes the point of classic stories where only some characters believe in the equivalence of, say, Oedipus' wife and mother. For Maida, they must all, in some sense, believe in it.

Another important contrast is with Kronfeld's (1990) notion of conversationally relevant descriptions, one which claims to account for why New York City can be variously described as "The Big Apple" or "The City with the World's Largest Jewish Community" but only in certain dialogue contexts. Our difference with him is that his explanation is entirely in terms of traditional distinctions like attributive-referential, and rests upon dubious notions like the non-contingency of generalizations such as "Anyone who was Smith's murderer must be insane" versus the extensionally equivalent (but merely contingent) "Anyone who was Joe's uncle must be insane." The problem here is that all the weight of the argument is shifted to the nature of particular generalizations, as was the case in the old analytic-synthetic dispute.

Someone who says Smith's murderer is insane may well not be doing so on the basis of sufficient evidence to allow the proof of that statement that Kronfeld's analysis requires. Nonetheless, the statement can be understood and properly linked to other descriptions. We prefer to concentrate on procedures that allow identifications to be made on the basis of certain environments' contents, a quite separate matter from the status of generalizations.

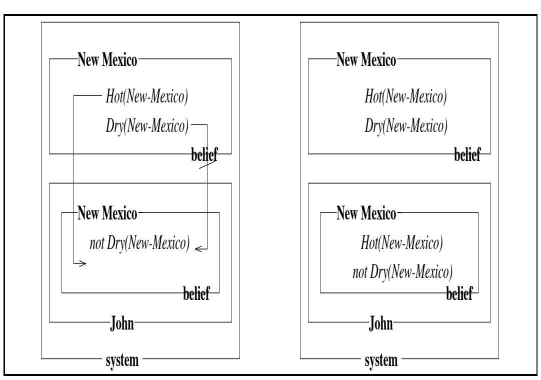
In our model, finding a referent in a conversation is the same as locating a topic belief environment. If a speaker thinks an addressee believes that two topics, say, the Morning Star and the Evening Star, are quite different, it should be possible for the speaker to intend to locate one (say, the Morning Star) for the addressee without referring to the other (the Evening Star), even when they are the same for the speaker. Conversely, one may well want to talk before certain audiences as though two topics are the same (e.g. the “languages” Serbian and Croatian) even when one believes they are different. Before we present the reference protocols in detail, we will first introduce the belief ascription system ViewGen which will enable us to represent the varying beliefs of agents.

ViewGen: the basic belief engine

A computational model of belief ascription is described in detail elsewhere (Ballim & Wilks, 1991) and is embodied in a prolog program called ViewGen. The basic algorithm of this model uses the notion of default reasoning to ascribe beliefs to other agents unless there is evidence to prevent the ascription. Perrault and Cohen (1981) also explored a belief and speech act logic based on a single explicit default axiom. As our previous work has shown for some years, the default ascription notion is basically correct, but the phenomena are more complex than are normally captured by an axiomatic approach. We offer no formal description here of the objects we propose and manipulate, and a reader for whom those (rather than the procedures and programs) are significant is referred to extensive formalization in (Ballim and Wilks, 1991).

ViewGen's belief space is divided into a number of explicit, topic-specific partitions, called *topic environments*. ViewGen also generates a type of environment known as a *viewpoint*. A viewpoint consists of some agent's belief set about some topics, parcelled up into topic environments. Within ViewGen, all beliefs are ultimately beliefs held by the system itself (e.g., the system's beliefs about France, what the system believes John believes about cars, etc.) and so, trivially, lie within the system's viewpoint. Part of the system's view of some topic (say, New Mexico) can be pictorially represented as in Figure 1.

Figure 1. The system's view of John’s view of New Mexico.



This diagram contains two types of environments: On the left, there is the box labeled with ``system'' at the bottom. This is a ``believer environment'' or ``viewpoint.'' Viewpoints contain topic environments, such as the box labeled with ``New Mexico'' at the top. A topic environment contains a group of propositions about the ``topic.'' So, for example, the above diagram conveys that the system believes that New Mexico is hot and dry and the lower box on the left shows that it also believes John believes New Mexico is dry.

Thus, if the topic of a topic environment is for a person, the topic environment may contain, in addition to the beliefs *about* the person---which would be in a topic box for John, with “John” at the top edge, and could include his height, say----- a viewpoint environment containing particular beliefs held by that person about various topics. Normally and for obvious reasons of efficiency, this is only done for those beliefs of a given person that are, as some would put it, *reportable*, where that will normally mean beliefs that conflict with those of the system itself. For example, suppose the system had beliefs about a person called John who believed that the Earth is flat, the systems John’s beliefs box for the Earth would show a belief inconsistent with the content of the system’s own “Earth” box, otherwise not..

Environments are dynamically created and altered. The basic algorithm of interest in this paper is an amalgamation mechanism that *ascribes* beliefs from one viewpoint to another (or, pictorially,``pushing one environment down into another''); ascribing certain beliefs, transforming some, and blocking the ascription of others. The simplest form of this algorithm is that a viewpoint should be generated using a default rule for ascription of beliefs. The default ascriptional rule is to assume that another person's view of a topic is the same as the system's own *except where there is explicit evidence to the contrary* in the form of a contrary belief already in the “receiving” or lower box. In examples of this sort, where the topic is also the agent into whose environment an ascription is being attempted, propositions in an outer topic environment (for the topic New Mexico, in the example above, in the rightmost box), are pushed inwards into a topic environment (for the same topic) within a believer viewpoint (John's) nested within the system. The result, shown on the above, is that the belief that New Mexico is hot is ascribed to John by the system in this way (indicated by the arrows on the left and the result on the right).

For the sake of convenience, in the body of the paper we will frequently call topic environments ‘boxes’ but use the word 'environment' for viewpoint environments. These nestings of environments are constructed on demand, nested as deeply as necessary for the processing in hand.

Some extensions and refinements to the basic program

In the tradition of Frege, topic environments in ViewGen can be classified into two different types: the topic environments (or boxes) for types and the topic environments (or boxes) for entities. Examples of topic environments for the latter are the person John or New Mexico in Figure 1. If the topic environment in Figure 1 had been "atom" about which John and the system had beliefs, then that would have been a topic as a type. All the topic environments in ViewGen are also treated as intensional objects and are, in that sense, concepts. Conventionally, we will write {…} round the box's topic name or description to denote the box or environment corresponding to the name or description, and at the base level of the system's own beliefs about the topic, unless indicated otherwise.

In (Ballim, Wilks & Barnden, 1991), a procedure was presented by which the system, which believes that two topic environments A and B are different, could construct a topic environment A-as-B specifically for another agent who believes that A and B are the same entity. Since this process can also be applied in the reverse direction, this convergence operation B-as-A can also be allowed within the system's own environment i.e. in a situation of a believer who believes two entities are the same but the other (modeled) believer believes them different. This move is consistent with the view that belief ascription is an updating procedure which ascribes from any previous environment to an updated environment. Hence the same convergence procedure can be performed when the system itself realizes that two topic environments are the same. The individual environments from which A-as-B was composed are then removed for the relevant believer.

So, different topics that are believed (by the system or belief environment host) to be the same should share the same topic box. For instance, *Morning_Star*, *Evening_Star* and *Venus* should share the same entity topic box in the system's environment if the system believes they are the same. The shared box will have a composite topic *Morning_Star-as-Evening_Star-as-Venus*, and this shared box can be regarded as the representation of an intensional object, namely, the system's own concept of that particular planet. However, it is possible for the system to realize that some other believer might think that the Morning Star, the Evening Star and Venus are different entities. So in the system's belief environment for that individual there would be three different environments.

ViewGen has a general assumption of box-parsimony as regards co-referentiality: we can indicate co-referentiality explicitly but we should assume that, within a single believer's overall space, a situation of multiple boxes for a single entity is an unstable one awaiting an amalgamation (akin to a higher level of garbage collection). Within a single believer's own space, differences of ontology cannot occur. All this means that we need to retain a co-referentiality notation for unstable situations.

This also means that tolerating co-referentiality as an unstable position (one with n coreferential boxes in the same believer's space) does not also mean that we must individualize boxes by every available description. But what must be explained is what can cause us to split a concept after provocation by another believer's ontological "errors" e.g. Smith thinks Jupiter is the Morning Star. As we mentioned earlier, similarly described topic environments in different belief environments also correspond to each other. More specifically, all those which share topic(s) will correspond to each other. For instance, both *Morning_Star* labeled boxes and *Evening_Star* labeled boxes will correspond to *Morning_Star-as-Evening_Star-as-Venus* labeled boxes but not to each other if they are in different belief environments. Conversely, if a new cross-environment correspondence is established, a new topic label needs to be dynamically created to represent this newly found correspondence relation.

If all parts of a topic are parts of another topic, we will say this second topic subsumes the first topic. Given a topic a, we will define its *reachable ascription set (RAS)* as the set of all the topic environments (including the ones that may be potentially generated by ascription) with a topic by which a is subsumed. It will be written as [a]. All the environments in [a] are reachable from the topic a by a process we will explain below. Notice all the environments in an RAS will correspond to each other. On other hand, some environments in one RAS might correspond to some environments in another RAS while the rest might not. This reflects the intransitive properties of the correspondence relation. Intuitively, the RAS of a topic includes all the topic environments in different viewpoint environments that are explicitly concerned with this topic.

An RAS is potentially infinite since it includes implicit boxes implied by the ascription process. But the recursive procedure of ascription provides a mechanism to produce any particular element of the set from the explicitly represented environments when necessary. Later below we will show how this recursive mechanism deals with this potentially infinite set in a tractable manner.

Cross-environment correspondence is commonly represented by specifying explicitly a relation between two entities which correspond (Maida, 1991) (Fauconnier, 1985). However, because of the possible disparity between

Different environments, the correspondence relation is nontransitive. So, to represent n objects in correspondence to each other, this method will require 2*n explicitly represented relations. In our representation, the correspondence is implied, and topic names and all the topic environments which share a topic will automatically correspond to each other, and the shared topic need only be repeated n times.

A model of reference

There are many more factors than reference involved in language: We need to define more clearly here what the problem is that our reference model is supposed to address.

There is strong similarity between a general human communication model and the structure of a computer network. Each communicator has layers of processing and a lower layer that provides services for a higher layer. Some ambiguities found at a lower layer may be passed to a higher layer to be resolved. Each layer has its own protocol. The protocol says not only what each system should do but also what another agent might do. For example, the speaker protocol would tell him how to construct a sentence so that the addressee can best figure out his intention. It will also tell him how an addressee might react to his utterance. Finally, it will tell him what he needs to do to be prepared for further interactions.

Each side of the communication, on this view, might have different protocols as long as they are compatible with the information being passed. This means we only need to consider how one system works without bothering whether it assumes correctly exactly what the other does. Reference protocols are "higher" than layers such as utterance parsing/generating but "lower" than layers such as planning. In this paper, we will only consider two types of sentences:

S --> the D P

S --> A believes S

where D is a definite description, P is a predicate about the referent indicated by D, and A is an agent. The representation of an utterance can be considered as a pair (D,P) and a believer A (or a list of believers in case of embedded belief reports) if the utterance is a belief report. To separate the reference layer from the higher layers, we note that the reference layer needs to answer two questions. The first is to locate a topic environment where the D and P from the utterance will be attributed. Secondly, after they are attributed, we ask where they can be further ascribed. The first problem is decided by a topic and a viewpoint environment. The second problem depends on the RASs the located environment belongs to. So the central problem here is locating a topic.

In our model, locating the referent of an utterance is assumed to be the same as locating a topic, a specific naming of an environment. The topic located indicates what kind of topic box the predicate will be attributed to. This topic will also guide where the asserted predicate (the one placed in the appropriate topic box) may be further ascribed. The topic (or candidate referent) can be either a simple or a composite topic. The information for topic location is given by the description part as well as the predicate part of the utterance.

When the system/speaker intends that the reference of an utterance is some intensional object, it would normally also believe that the other party to the conversation has an intensional object that corresponds to it. The system refers only in terms of its model of the other's belief, and not what the other may actually believe. Even the intensional objects that correspond to each other can be drastically different in fact (whatever that may be taken to mean here) and they may also be believed by the system not to correspond to any object at all in the real world.

We believe, as we noted earlier, that any reference theory (e.g. Maida, 1991) that hypothesizes a common external world (beyond belief environments) cannot deal properly with the situations where the entities the speaker believes in do not match very well what the addressee believes in. For example, suppose John, who believes Morning Star is the Evening Star, wants Mary, who believes the Morning Star is different from the Evening Star, to believe that the Morning Star symbolizes good luck while the Evening Star symbolizes bad luck. He must be able to talk to Mary about the Morning Star being so and so and the Evening Star being some other so and so. Insisting that the references given by John all refer to the same planet is of no help in understanding this conversation.

The goal of the rest of this paper is to try to spell out in very primitive terms the reference protocol implied by these assumptions and how it will deal with some traditional issues. The protocol routine is concerned with the following problems:

1) How to find a referent (or reference topic) from a description D or, conversely, given a referent, how to choose a description for it?

2) What belief environments are relevant to an utterance?

3) Given a referent and relevant belief environments how and where will the predicate P be ascribed?

Question (2) is relatively simple. The relevant belief environments are the belief environments the system needs to consider when producing or hearing an utterance. If the utterance is a belief report, the belief environments which belong to the agent whose beliefs are reported are relevant. Otherwise, the system's and the other communicator's environments are relevant. Questions (1) and (3) are addressed in the next sections.

Descriptions

The description D in the definite noun phrase "the D" is produced by a speaker so that the addressee can pick out a referent with close relations to what D is intended to apply to. To accomplish this, the speaker needs to consider whether and how the addressee can pick out a referent with D. So, he should believe D will describe some referent for the addressee. Conversely, the addressee needs to consider what the speaker intends, and believes D will describe the referent for the speaker. The complexity of the problem is due to the recursiveness of this process. Clark & Marshall (1981) tried to show that if any part of the following-----

SB ~D(r)
SB HB ~D(r)
SB HB SB ~D(r)
SB HB SB HB ~D(r)…. ad infinitum,
where SB stands for "Speaker believes that" and HB stands for "Hearer believes that"-----

is true about the description D, the speaker should avoid using D as a referring expression since the referring, they argued, will not be successful if what he believes is true. They proposed a condition on the use of the description D which involved their so-called mutual belief condition. Using Donnellan's famous martini example, Perrault and Cohen (1981) showed that the Clark and Marshall condition is too strong and that, as long as the corresponding mutual belief is believed to hold by one of conversation's participants, the reference can be successful, even when the speaker and/or the hearer in fact has different private beliefs. Formally, they expressed their condition as:

(c1) A description D can be used by the speaker(s) for referring referent r (for hearer H) only if (where MB is "mutually believes"):

MB(S,H) D(r) or SBMB(H,S) D(r) or SBHBMB(S,H) D(r) or ....

However, Perrault and Cohen's condition is still too strong: as long as both the speaker and the hearer believe one of them believes D uniquely describes referent r (even when they both believe the other does not think so), D can be successfully used to pick referent r. Let us redefine the (uncomputable) MB condition listed in (c1) as MB'(S,H) D(r), which is to unpacked as: SBHB D(r) or SBHBSB D(r) which is compatible with both S~B D(r) and H~B D(r), where both of those are from the point of view of the system itself, which we are taking as a form of "ground truth". Using this new (finite) condition MB' we can then replace (c1) with:
(c2) A description D can be used by the speaker for referring referent r if

MB'(S,H) D(r) or MB'(H,S)SB D(r) or MB'(S,H)HB D(r) and/or MB'(S,H)TB D(r) or MB'(H,S)SBTB D(r) or MB'(S,H)HBTB D(r)
iin the latter cases if a third party (T) is involved (which will be John in the example below, who is neither speaker nor addressee).

These conditions might seem to suggest that an infinite number of recursively embedded beliefs need to be explicitly represented and known to be true. However, this cannot be the case. It is counter-intuitive and computationally impossible to keep track of an infinite number of such beliefs. Also, for Clark & Marshall--- as for Perrault & Cohen---- the referent is in a common world shared by all the agents. However we can use the same notation here but with a slightly different meaning, relativised to a situation with no objective truth outside the participants. For us, r is a topic and D(r) is true in an environment if there is a topic environment in which we can find D(r) or D(r') where r' is a composite topic of which r is a part (this sia notion we have no space to explain here).

As we have shown elsewhere (Wilks, 1982), we can in this way detach mutual belief from its originating, but highly restrictive, notion of physical co-presence (Clark & Marshall, 1981) and demonstrate that all the benefits of mutual belief are inherent in that of ascribed belief, with the ascription process actually being carried out as many times as resources permit, or the example requires, but with no need to set up any kind of infinite recursions to explain a concept of mutual belief. We will then define an *ascription path (AP)* in an RAS as the set of topic environments that can (or potentially can) be produced by such a sequence of ascriptions into a sequence of embedded viewpoint environments in the RAS.

So, we can now relativize "mutual belief" to the ViewGen system S and one agent H as a belief D (r) such that SB D(r) and there are no explicit beliefs in any of the belief environments listed in (c2) that will prevent it being ascribed to that environment. The conditions listed are only implied by ascription, and ascription is a finite and computable process. All the possible blocking beliefs preventing a particular belief being ascribed must be relevant to it. Secondly, all the beliefs relevant to a belief can be retrieved in finite time since only finite number of explicit beliefs can be stored.

Therefore, whether there are "blocking" beliefs for a given belief can also be decided in a finite time. This process is similar to that of mathematical induction where finite recursive proofs are given to infinite sets. It is in this procedural, ViewGen-related, sense that we shall continue to use the MB predicate in this paper.

Now we can go back to the problem of descriptions used for referring.

If we define MB"(S,H,A)D(r) as:

MB'(S,H) D(r) or MB'(H,S)SB D(r) or MB'(S,H)HB D(r) or MB'(S,H)ABD(r) or MB'(H,S)SBABD(r) or MB'(S,H)HBAB D(r) (where A is the agent whose belief is reported assuming the utterance is a belief report, otherwise A is null).

Condition (c2) should now be restated as:

(c2') A description D can be used by the speaker for referring referent r if D(r) is true in one of the belief environments and there is no explicit belief that can prevent it being ascribed indefinitely along [r] so that MB"(S,H,A)D(r) would eventually be true.

The condition (c2') is from the speaker's point of view. From the addressee's point of view, we get a symmetric condition. Following the discussion above, both can be checked in a finite time. We can now summarize the addressee protocol procedurally as follows (and the **speaker protocol** is symmetrical):

The addressee protocol (as inferred by the speaker-system):

1. build a working box with a variable topic X which contains both D and P.
2. find a reference topic r so that the descriptions Q in the working box can be found in the system's belief space and ascribing them along an ascription path in [r] will make MB"(S,H,A)Q(r) eventually true. If everything else is equal, a less specific topic is preferred.
3. instantiate X of the working box as r.
4. find a topic box in a relevant belief environment which has a topic with r as its part and which is closest to the system environment such that most of the working box can be ascribed into the topic box and ascribing the ascribed part Q along an ascription path in [r] further will cause MB"(S,H,A)Q(r) eventually to be true.
5. perform the ascription decided in 4.

Steps 2 and 4 are just a variation of condition (c2"). They can all be performed in a finite time. At step 2 and 3 rules of inheritance and explication may need to be applied on the working box and the topic environment to be matched in order to satisfy (c2'). Finally the algorithm presented here is designed for the purpose of conceptual clarity. It is not the most efficient algorithm.

Reviewing traditional issues using this model: *de dicto* vs. *de re* belief reports

One traditional distinction between belief reports is between *de dicto* and *de re* beliefs. For instance in:

Ralph believes the janitor is a spy.

what Ralph believes *de dicto* is what Ralph would explicitly assent to on the basis of what he believes, i.e. if Ralph were asked about the janitor and his possibly being a spy Whereas, what Ralph believes *de re* means that he believes the person referred to is a spy even though he may not know that he is a janitor. To decide whether a belief report is *de dicto* is to check whether the agent whose belief is reported would agree with what the reference description says about the referent. If he would agree then it is *de dicto*. Otherwise it is not. Similarly, if the speaker and the addressee believe the description is true for the referent it is *de re* (unless there is evidence they are both deceived with respect to some other belief space that is definitive, such as the system's own space) otherwise it is not. Since the parties involved may all believe the description is true for the referent, a belief report can be both *de dicto* and *de re*. In our model, referent location depends on how (c2') is satisfied. Condition (c2') indicates that in a belief report the referent can be identified by either what the speaker believes he and the addressee believe or what the speaker believes he and the addressee believe the agent (whose belief is reported) believes. The former corresponds to *de re* and the later corresponds to *de dicto*. Sometimes both will give the same result. Our point is that the traditional distinction can better be seen as one between two sub-procedures. The procedural distinction can then be held onto even when the traditional metaphysical distinction becomes unclear; to see this consider the utterance:

John believes the winner is a mediocre player.

Suppose the system is the addressee and holds the following a beliefs before that utterance (where this state is taken to be one *without* the presence of the key belief $$ below). First, we will build what we might call a "working box" from the sentence (step 1): one containing beliefs Winner (X) and Mediocre_player (X). In step 2 the description *winner* in the system's environment will allow the working box to locate *Federer* as a referent which will in turn instantiate the variable X in the working box (step 3).

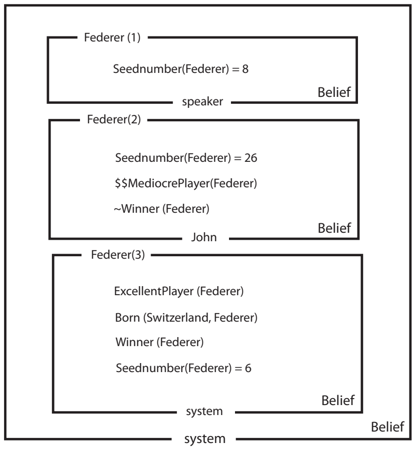


Figure 2. System's beliefs before and after (with $$) the utterance on the winner.

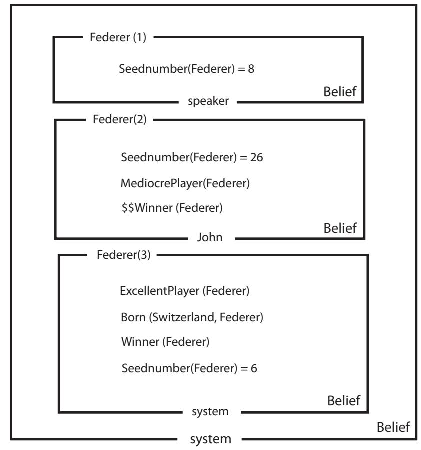


Figure 3. System's beliefs before and after (with $$) before the utterance on the winner, with John's prior state changed.

Since John's belief environment is the relevant belief space and box (2) is the closest relevant topic environment when we ascribe the instantiated working box into (2), we will find that *Mediocre_player(Federer)* will be ascribed and it will satisfy the condition (c2"). However, since there is a description *~Winner* in that box *Winner*(*Federer)* cannot be ascribed by step 2 into what John believes i.e. in (Federer(2)). Therefore the utterance is *de re* but not *de dicto*, and what the system believes after the utterance is as shown in Figure 2 with the belief marked $$ in place.

If we now look at the slightly different situation in Figure 3, where there is no *~Winner(Federer)* in John's beliefs about Federer. The difference in this case is that in this case the description *Winner* can be ascribed from box (3) to box (2). That means the referent can now be matched in step 2 both by what the speaker believes and by what John believes. So in this case the belief report is both *de re* and *de dicto*. Based on this process, we can see that if it is agent A's belief that is reported, and there are no explicit beliefs in A's belief environment that will prevent the description being true, we will assume the report is also *de dicto*. In yet another situation, the system might believe that Federer is not the winner but believe that John believes he is. From condition (c2'), the referent that will be matched is still Federer, this time using A, the third-party condition, and so the first ascription will still be made. Such a situation is hard to classify as either *de dicto* or *de re*, yet it is a perfectly natural one in belief ascription terms. Wiebe (1990) has showed such usage can be frequently found in real discourses.

Conclusion

These examples present problems of classification for the traditional *de dicto/de re* distinction, and suggest that, since the reference will be resolved by the same procedure in all the cases, the traditional distinction serves no particular purpose, procedurally or as a classification of examples. This account has led us to a purely procedural account of time-honored distinctions like de *dicto/de re*, which lose much of their appeal when the belief differences of individuals are taken seriously.

References

Ballim, A., and Wilks, Y. 1991. *Artificial Believers*. Hillsdale, NJ: Erlbaum.

Clark, H. and Marshall, C. 1981. Definite reference and mutual knowledge. In eds. Joshi, Webber and Sag. *Elements of Discourse Understanding.* Cambridge: Cambridge University Press.

Fauconnier, G. 1985. *Mental Spaces*. Cambridge, MA: MIT Press.

Kronfeld, A. 1990. *Reference and Computation*. New York: Cambridge University Press.

Maida, A. 1991. Maintaining mental models of agents who have existential misconceptions. *Artificial Intelligence,* 50. pp 331-383.

Perraut, R. and Cohen, P. 1981. It's for your own good: a note on inaccurate reference. In eds. Joshi, Webber and Sag. *Elements of Discourse Understanding.* Cambridge: Cambridge University Press.

Wiebe, J. 1990. Recognizing subjective sentences. SUNY Department of Computer Science, Buffalo, NY.

Wilks, Y. 1982. Comments on Clark and Carlson's paper. In ed. Smith. *Mutual Knowledge*. New York: Academic Press.

Wilks, Y. Wilks, Y., Barnden, J. and Wang, J. 1991. Your metaphor or mine? Belief ascription and metaphor interpretation. In *Proc. 12th IJCAI.* Los Altos, CA: Morgan Kaufman.

Wilks, Y., Worgan, S., Dingli, A., Catizone, R., Moore, R., Field, D. and Cheng, W. 2011 A prototype for a conversational Companion for reminiscing about images. *Computer Speech and Language*. pp. 221-240.