Intentional identity
*From Hob and Nob to Arsky and Barsky*

van Rooij, R.

**DOI**
10.1002/9781118788516.sem128

**Publication date**
2021

**Document Version**
Submitted manuscript

**Published in**
The Wiley Blackwell Companion to Semantics

**Link to publication**

**Citation for published version (APA):**

**General rights**
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
1 Introduction

You and I might have different beliefs about president Trump, but our beliefs have a common focus: Donald Trump. Sometimes, however, multiple agents can have attitudes with a common focus, even though there is nothing at that focus. Hob and Nob, for instance, can have fears about the same witch. It is an interesting philosophical puzzle how to account for the latter type of common focus. It is what Geach dubbed a problem involving intentional identity:

We have intentional identity when a number of people, or one person on different occasions, have attitudes with a common focus, whether or not there actually is something at that focus. (Geach, 1967, p. 147)

In this paper we are primarily interesting in the linguistic manifestation of this problem, a problem involving the analysis of anaphora. We will see that the linguistic manifestation of the problem is closely related with Geach’s original philosophical problem.

According to a prominent (though not generally accepted) view in semantics, so-called unbound pronouns – pronouns not bound by a quantifier $Q$ inside the smallest clause containing $Q$ – should be treated either as abbreviations for the antecedent clause (see entry on E-type pronouns) or as variables bound by a (dynamic) existential quantifier (see entry on Donkey anaphora). Geach’s (1967) notorious Hob-Nob sentences, exemplifying intentional identity attributions, have always been a threat to this assumption.

In this paper, I relate the problem that Geach’s Hob-Nob sentences pose for the traditional analyses of pronouns to the problem that examples of pronominal contradiction pose for the same theories. First, I will point out both the similarities and the differences between anaphora across attitude clauses (or Hob-Nob sentences as I will call them), on the one hand, and cross-speaker anaphora or Hob-Nob situations), on the other.

In Section 2 of this paper, I will discuss Geach’s traditional problem of intentional identity. In Section 3, I will describe Edelberg’s (1986) more recent asymmetry problem. In section 4 a de re analysis of intentional identity attributions is discussed, making use of counterpart theory. In section 5 I will consider how to address Edelberg’s asymmetry problem following standard approaches towards anaphoric dependence, by relating anaphora
across attitude clauses to cases of cross-speaker anaphora. In Section 6, however, I show why these obvious proposals won’t work in general. I will suggest that these problems can be solved, however, if we take the notion of speaker’s reference (see entry on Speaker’s reference) seriously, as I discuss in Section 7. Section 8 concludes this paper.

2 The problem of intentional identity

A key problem that every semantic account of anaphora faces is that a pronoun occurring in the embedded clause of an attitude attribution can have as its syntactic antecedent an indefinite in the embedded clause of an earlier attitude attribution. In a logical language this is not difficult to represent if the indefinite is interpreted *de re* (see entry on Attitude reports de dicto and de re). But the problem is that this doesn’t always seem to be the case. This is the linguistic manifestation of the problem discussed under the heading of *intentional identity* by Geach (1967), and known as the unsolved problem of *de dicto pronouns* by those working in the tradition of Montague semantics (Montague, 1973).3 Examples of these sentences include the following:

(1) Bill believes that a woman stole his car.
   He also believes that she is now hiding from the police.

(2) Obelix wants to catch a boar today, and he wants to eat it afterwards.

(3) Hob believes that a witch blighted Bob’s mare,
    and Nob believes she killed Cob’s sow.

On the intended readings of these sentences, the attitude attributions can be true without there being a woman about whom Bill has beliefs, a boar that Obelix wants to catch, or a witch that is responsible for the beliefs of Hob and Nob. For (3) to be true, there does not even have to be an existing individual that is the focus of both Hob’s and Nob’s beliefs. This is shown by the following Geachian story:

Last night, Bob’s mare became quite ill. Hob, who tends Bob’s barn, inferred that a witch blighted her. This morning Hob said to his friend, Nob, “A witch blighted Bob’s mare.” Nob believes what Hob has told him. He thinks for a moment, and says, “Cob’s sow died early this morning. I’ll bet the same witch killed the sow, too.” But in fact both animals fell ill due to perfectly natural causes. (Edelberg, 1986, pp. 1-2)

According to this story, the Hob-Nob sentence (3) would be true. In the Geachian tradition, anaphoric elements are treated as bound variables; but the problem is that there is no way to bind the variable that represents the pronoun in the second clause by the quantifier that represents the indefinite in the first clause if you can quantify only over existing individuals.
In the framework of traditional Montague semantics, the following translations would come out naturally (where $h$ stands for ‘Hob’, $n$ for ‘Nob’, $W$ for ‘witch’, $BBM$ for ‘blighted Bob’s mare’, and $KCS$ for ‘killed Cob’s sow’):

(4) a. $\text{Bel}(h, \exists x [Wx \land BBMx]) \land \text{Bel}(n, KCSx)$  
    b. $\text{Bel}(h, \exists x [Wx \land BBMx \land \text{Bel}(n, KCSx)])$  
    c. $\exists x [Wx \land \text{Bel}(h, BBMx) \land \text{Bel}(n, KCSx)]$

In (4a), the pronoun is not treated as a bound variable, and cannot be interpreted. If pronouns are treated as bound variables, it seems that the only possible way to go is to use either representation (4b) or (4c). Unfortunately, (4b) doesn’t give the intended reading because the attitude attribution doesn’t seem to say anything about what Hob believes about Nob’s beliefs, and representation (4c) does not predict (3) to be true in the above story because, in fact, witches do not exist.\footnote{4}

From these problems, one might conclude that variable $x$ should really range over non-actual objects—objects that exist in other possible worlds but not in the actual world—and that cases of intentional identity should be translated as in (4c) after all. According to this view, in cases of intentional identity, a de re belief attribution is made about a specific object that need not exist in the actual world. Under the influence of Quine’s attack in the 1950s and 1960s, the idea that we can talk about objects that don’t exist in the actual world was for a long time extremely unpopular in analytic philosophy. Due to the work of Parsons (1974), Priest (2005) and others, this view — which according to Priest was the dominant view in Ancient and Medieval logic — has become somewhat more respectable now. Indeed, we do talk about Farther Christmas and Gandalf, although they actually do not exist. One worry with an analysis of (3) as (4c), however, is that a seemingly similar sentence like (2) doesn’t seem to be about a specific (perhaps non-actual) boar at all. There does not need to be one specific boar that Obelix’s belief is about such that Obelix believes he will catch it and wants to eat it afterwards to make the attitude attribution true. Given the unclear identity conditions of non-actual objects, it is not obvious how problematic this issue really is, however. More seriously, for (3), for instance, to be true, it should be predicted that in all of Hob’s belief alternatives there is a witch who blighted Bob’s mare, something that is not guaranteed if we represent (3) by (4c).

The last problem suggests that we should follow Slater (1988) and represent intentional identity attributions in a non-Montagovian way, by splitting the existential from the descriptive contribution of the indefinite ‘a witch’, as in (5), and so still be able to provide a de re analysis of Hob-Nob sentences:

(5) $\exists x \text{Bel}(h, W(x) \land BBM(x)) \land \text{Bel}(n, KCS(x))$

According to it, Hob and Nob have a belief about a specific object, but all we know about this object is that Hob thinks that it is a witch that blighted Bob’s mare and Nob believes
that it killed Cob’s sow. Unfortunately, (3) can intuitively be true without any specific object satisfying the above conditions. The reason is that there need not be one actually-existing object that is responsible for the relevant beliefs of Hob and Nob. Hob believes none of the individuals he has ever come across to be a witch; thus none of them satisfies the property expressed by \(\lambda x.\text{Bel}(h, W(x) \land BBM(x))\) (cf. Buridan, 1350).

All of these problems seem to suggest that we should indeed represent a sentence like (3) by (5), but that the variables should range not over specific actual objects, but either over possible objects, or individual concepts instead. In fact, Priest’s (2005, 65, n. 12) proposal boils down to the first of these suggested proposals. He assumes that ‘\(\exists xPx\)’ should read as ‘Something, \(x\), such that \(Px\)’, rather than ‘There exists something, \(x\), such that \(Px\)’. A proposal along the the lines of the second suggestion was made by Saarinen (1978) to account for intentional identity attributions. He assumed that variables range over (non-rigid) individual concepts and that these concepts don’t have to be instantiated in the actual world.

There are problems with both of these suggestions. First, an analysis in terms of unrestricted individual concepts is clearly too weak. If there is no limitation on the individual concepts quantified over, the analysis can hardly explain why intentional identity attributions are about the same object, or better, why they have a common focus.

To solve this problem, it has been suggested (e.g. van Rooij & Zimmermann, 1996; Zimmermann, 1999, Perry, 2001) that for the multiple agents to have a common focus, it is required that their attitudes are linked together by a causal chain. The beliefs of Hob and Nob have a common focus in our Geachian story, because both of their beliefs are causally linked via a conversation. In a similar Geachian situation, both beliefs have a common focus because they are caused by the same newspaper article. Unfortunately, there exist convincing examples showing that a causal link is not, in general, necessary for intentional identity attributions to be appropriate (cf. Edelberg, 1992; van Rooy, 2000; Sandgren, 2016).

An analysis along the lines of Priest does not face Saarinen’s problem: both beliefs can just be about the same, non-existing, object, if one assumes—as Priest (2005) does—that the domain that is quantified over is the same in each world, even though not all elements of the domain need to correspond with existing individuals in each world. Unfortunately, Priest’s proposal, just like Saarinen’s, faces another difficulty: Edelberg’s asymmetry problem.

3 Edelberg’s Asymmetry problem

If we don’t restrict the range of the variables, Saarinen’s and Priest’s proposal would predict that attributions of the form (6) are equivalent to attributions of the form (7):

\[(6) \exists x\text{Bel}(h, Px) \land \text{Bel}(n, Qx)\]
However, Edelberg (1986, 1992, 1995) observed that intentional identity attributions are in general not symmetric. Consider the following case of Sandgren (2016, p. 49), inspired by Edelberg’s (1995) case of Arsky and Barsky:

Hob and Nob are partners in a witch-hunting business. They investigate the suspicious illness of Bob’s mare. They both conclude that Bob’s mare was blighted by a witch who was working alone. The next day they investigate the suspicious death of Cob’s sow. At this stage a disagreement arises between Hob and Nob. Nob thinks one witch both blighted Bob’s mare and killed Cob’s sow. Hob think one witch blighted Bob’s mare and a different witch killed Cob’s sow. Witches do not exist.

For this case we find (8) but not (9) acceptable:

(8) Hob thinks a witch has blighted Bob’s mare, and Nob believes she (the same witch) killed Cob’s sow.

(9) Nob think a witch has killed Cob’s sow, and Hob believes she (the same witch) blighted Bob’s mare.

Intentional identity attributions, then, are in general not symmetric. Edelberg called this problem the asymmetry problem about intentional identity. Any proposal that seeks to account for intentional identity by representing sentences like (3) by (5) and by allowing unrestricted quantification over possible objects or individual concepts fails to explain this asymmetry.\(^6\)

A different but related problem is discussed by Edelberg under the heading of the variable aboutness problem of attitudes de re. The problem is related to the following case (again modeled after an example due to Edelberg, 1995):

Bob’s mare and Cob’s sow are dead. A single person killed both of them. Hob investigates both killings, and comes to believe that someone killed Bob’s mare and that someone killed Cob’s sow, but he doesn’t have anyone in particular in mind as a suspect. Hob does not believe that the one who killed Bob’s mare is the same person as the one who killed Cob’s sow.

The problem is to account for the intuition that on their most straightforward readings, (10) is true, while (11) is false:

(10) Someone killed Bob’s mare, and Hob thinks he didn’t kill Cob’s sow.\(^1\)

(11) Someone killed Bob’s mare, and Hob thinks he killed Cob’s sow
The problem for an approach on which variables range over concepts is that such an approach predicts that (11) as well as (10) is true, because there is a single concept, the one who killed Cob’s sow, whose instantiation in the actual world killed Bob’s mare and whose instantiation in Hob’s belief worlds also killed Cob’s sow in each of them.

Now we have three kinds of problems. First, we have cases like (1) and (2), where only one agent is involved and the pronoun in the second sentence does not refer back to a specific existing object that the speaker refers to. Second, we have *de re* attributions like (10) and (11), where the pronoun in the second sentence *does* refer back to such a specific existing object. And third, we have intentional identity attributions like (3), where two agents are involved and the pronoun does not refer, for the speaker, to a specific existing individual. For *de re* attributions we have to account for the variable aboutness problem; and for intentional identity attributions with more agents involved, we have to account for the asymmetry problem.

4 A *de re* analysis with contextual counterparts?

Although a *de re* analysis of Hob-Nob sentences where quantifiers range over all individual concepts is clearly too weak, perhaps such an analysis can be appropriate after all, if we limit the domain over which variables range. One way to account for this is by making use of counterpart theory.

Whereas Quine thought that modal statements about particular objects are not possible, he admitted that belief attributions about particular objects can be made. But he also showed that the most obvious way this can be accounted for within intensional logic would lead to inconsistencies. And indeed, the assumption that for a belief attribution about a particular individual it is always this individual that is referred to in all worlds compatible with what the agent believes seems to lead to embarrassing results.

Consider Quine’s (1956) Ralph who, one evening, sees a man with a brown hat whose suspicious behavior leads Ralph to believe that the man is a spy. On another occasion, Ralph sees the same man at the beach, but he does not recognize him as the same man; and the thought that the man he sees at the beach is a spy does not even occur to him. Intuitively, we can attribute his beliefs by saying (12) and (13):

(12) Ralph believes of the man with the brown hat that he is a spy
(13) Ralph doesn’t believe of the man he saw at the beach that he is a spy

But now the story goes on. In fact, the man with the hat who is later seen at the beach happens to be Ortcutt. So we seem to be allowed to infer (14) from (12), and (15) from (13):

(14) Ralph believes of Ortcutt that he is a spy.

---

1I am making use of this rather awkward phrasing to keep scope matters clear.
(15) Ralph doesn’t believe of Ortcutt that he is a spy.

Now, does Ralph believe that Ortcutt is a spy or not? Or better, how can we account for the beliefs attributed to Ralph that seem to be about Ortcutt without concluding that Ralph is irrational?

To account for belief attributions one can make use of counterpart theory. To determine which beliefs Ralph has about $d$ in $w$, we don’t look at which properties $d$ itself has in other worlds, but rather at the properties the counterpart(s) of $d$ has in these other worlds. Counterpart theory can be formalized in several ways. On Lewis’s (1968) formalization, a formula like $\Diamond Rab$ would be true in a world if in every world containing counterparts of $a$ and $b$, every counterpart of $a$ bears the relation $R$ to every counterpart of $b$. Alternatively, Stalnaker (1987) proposed that we should count a formula like $\Diamond Rab$ only true in a world if it holds that for every world in which $a$ and $b$ both have a counterpart, the counterpart of $a$ bears the relation $R$ to the counterpart of $b$. Thus, we should not quantify over counterparts, but interpret sentences with respect to a function that picks out a unique individual for each individual and world to be the counterpart, or representative, of that individual in that world.

If $c$ is such a counterpart function, $c$ is a function that takes an individual, $d$, and a world, $w$, as arguments, and has an individual in the domain of $w$ as its value. This latter individual might be called the counterpart of $d$ in $w$ with respect to $c$, $c_w(d)$. Formally this means that sentences are not only interpreted with respect to a world and an assignment function, but also with respect to a counterpart function. If we assume for the moment that all individual terms are either bound variables or proper names, atomic formulas of the form $P(x)$ are interpreted as $c_w(g(x)) \in I_w(P)$, while the belief attribution $\text{Bel}(t,A)$ is taken to be true iff $\forall v \in K([t])^{w,g,c}, A$ is true with respect to $v, g$ and $c$, where $K(a,w)$ denotes the set of worlds compatible with what $a$ believes in $w$.

Notice that according to the above interpretation, belief attributions are interpreted with respect to a single counterpart function, that assigns to each individual and world at most a single counterpart of the individual in this world. But if in some of Ralph’s belief worlds the actual Ortcutt has two counterparts, which one do we refer to by a belief attribution like Ralph believes that Ortcutt is a spy? According to von Stechow (1984) and Stalnaker (1988), among others, which representation or counterpart we refer to depends not so much on the belief state of the agent itself, as on the intention of the speaker and on the conversational situation in which the belief attribution is made. The counterpart/representative we refer to depends on the issue which one of Ralph’s representations of Oscar is most salient in the relevant conversational situation. Formally this means that pragmatically speaking there will be a unique most salient counterpart function with respect to which the sentence should be interpreted.

Notice that when we assume that it depends on context which representation is relevant for the analysis of the de re belief attribution, there might be conversational situations where we might truly say that Ralph doesn’t believe of Ortcutt that he is a pillar of society,
which indeed seems to be in agreements with the facts. If I only had given half of the story, and only told you that Ralph saw a man with a brown hat who behaves suspiciously, the belief attribution seems to be true.\(^7\)

One might now account for Hob-Nob sentences by representing them as in (5), repeated as (16):\(^8\)

\[(16) \exists x \text{Bel}(h, W(x)) \land \text{BBM}(x) \land \text{Bel}(n, KCS(x))\]

Now we can assume that there is one, perhaps non-actually existing object (call her ‘Samantha’) such that the unique conversationally relevant counterpart function picks out Samantha as the witch who blighted Bob’s mare in all of Hob’s doxastically accessible worlds, and Samantha as the witch who killed Cob’s sow in all of Nob’s doxastically accessible worlds, so that the sentence is true. To account for the asymmetry between (8) and (9) involving Hob and Nob, on the other hand, we might assume that the default conversational relevant counterpart function will be such that the function from worlds to objects that is associated with the one who (according to Hob) killed Bob’s mare simply corresponds with the description ‘the killer of Bob’s mare’ by which the ‘object’ is introduced into the conversation. One can hypothesize that it is only in special cases (like in Edelberg’s Samantha-case, to be discussed later) that this default assumption is overruled.

5 Asymmetry explained by descriptive approaches

Although an analysis of Hob-Nob sentences as context-dependent de re-attributions is not unnatural, it comes with at least one—in my eyes serious—drawback: To turn Hob-Nob sentences into de re-attributions, we have to assume a somewhat unnatural logical representation of the sentences by splitting the existential from the descriptive contribution of the indefinite ‘a witch’. That is, it seems most natural to represent a Hob-Nob sentence like (4a), repeated here as (17):

\[(17) \text{Bel}(h, \exists x [W(x) \land \text{BBM}(x)]) \land \text{Bel}(n, KCS(x))\]

In fact, an analogue representation seems definitely asked for to account for examples like (1) and (2), where only one agent is involved, and the pronoun in the second sentence does not refer back to a specific existing object that the speaker refers to. As we have seen, however, the problem with a representation like (17) was that the pronoun could not be interpreted if making use of standard first-order modal logic. But semanticists faced a very similar problem in the 1970s to account for anaphoric dependencies across sentential boundaries, even without the involvement of intensional attitudes. Fortunately, nowadays there exist several frameworks that can handle such anaphoric dependencies. It is only to be expected that the intentional identity cases discussed above could be handled in one of these frameworks, too. In fact, this is what I believe. But as we will see, the solution to these intentional identity problems is not as straightforward as one might hope.
5.1 E-type pronouns and dynamic semantics

The two most prominent frameworks that can handle anaphoric dependencies across sentential boundaries are the E-type approach (see entry on E-type pronouns), on the one hand, and dynamic semantics, on the other (see entry on Donkey anaphora). According to the E-type approach (Evans, 1977, Neale 1990), the meaning, or referent of a pronoun is determined by the definite or universal description recoverable from its antecedent clause. Thus, in a sequence of the form Some S are P. They are Q, the denotation/reference of the pronoun they is determined by the description (all) the S such that P. In sequences involving a singular pronoun of the form There is an S that is P. She is Q that we will mainly be concerned with in this paper, the description The S that is P is taken to be satisfied by exactly one individual.

Standard dynamic semantics, be it Kamp’s (1981) Discourse Representation Theory (DRT), Heim’s (1982) File Change Semantics (FCS), or Groenendijk & Stokhof’s (1991) Dynamic Predicate Logic (DPL), has become a very popular way to account for anaphoric relations across sentential boundaries. According to this theory, each sentential clause/formula is interpreted with respect to a unique context, where this context represents information about the subject matter of conversation and the values of variables. Whereas in traditional semantic theories the primary goal was to determine the truth conditions of sentences in a systematic way, in these more recent theories more attention is paid to the ways in which sentences change the context of interpretation.

In dynamic semantics, contexts are typically represented by sets of world-assignment pairs. In this way, a context can represent not only the ‘world’ information about what is presupposed with respect to the subject matter of conversation, but also the information about the possible values of variables, or discourse referents. If we fix a world, and concentrate only on the latter kind of information, we can represent a context by a set of (partial) assignment functions. An indexed sentence like

(18) A man\(_x\) is walking in the park

will now update a context, C, by enriching the assignments of this context; each new assignment will also assign a value to variable (or discourse referent) \(x\), and each man who is walking in the park in this fixed world will be the value of variable \(x\) with respect to one of the assignments of the updated context. Thus, the only information associated with \(x\) in this new context, \(C'\), is that the value of \(x\) is a man who is walking in the park in this world. As a result, a subsequent sentence like

(19) He\(_x\) is whistling

can now be interpreted with respect to this updated context. If we again fix a world, sentence \(A\) according to the above theories, will be true in this world with respect to assignment \(g\) iff the update of context \(\{g\}\) with \(A\), \([[A]](\{g\})\), is non-empty. Similarly, the
discourse $A_1, ..., A_n$ will be true with respect to assignment $g$ iff $[[A_n]](...([[A_1]](\{g\}))...)$ is non-empty. As a result, the discourse (18) - (19) is predicted to be true iff there is a man who is walking in the park and whistling. In general, it is predicted that $\exists x\phi \land \psi$ is equivalent with $\exists x(\phi \land \psi)$ (see entry on Donkey anaphora).

Note that these theories predict that in our above sequence the pronoun $he$ relates back to the (context-dependent) referent of the indefinite description a man who is walking in the park, which is recoverable from the antecedent clause. The reason is that the only information associated with variable or discourse referent $x$ in the context resulting from the update of the first sentence is that the value of $x$ is a man who is walking in the park.

5.2 Cross-speaker anaphora

Although dynamic semantics has been developed to account for anaphoric and presuppositional dependencies in discourses made by a single speaker, it seems we can also apply it to cases in which two or more agents exchange information about an object. Following Dekker & van Rooy (1998), I will call such examples Hob-Nob situations. Typically these involve the use of pronouns by one agent to refer back to objects mentioned or introduced by another agent — hence the term “cross-speaker anaphora”, which is also used. Consider the following dialogue between Hob and Nob:

(20) Hob: Someone killed Bob’s mare.

Nob: He also killed Cob’s sow.

Standard dynamic semantics seems able to account reasonably well for what is going on in these sentences. First, it seems obvious that for both sentences to be true, it has to be the case that someone who killed Bob’s mare also killed Cob’s sow, just as the above definition of the truth conditions of discourses predicts. Thus, if we want to be able to determine the truth conditions of the second sentence relatively independent of what is asserted by Hob we can treat the pronoun $he$ as the (context-dependent) referent of the indefinite description someone who killed Bob’s mare.

Second, it seems reasonable to make the Gricean assumption that if somebody makes an assertion, he should also believe what he asserts. But in normal cases there seems to be an asymmetry between what Hob has to believe and what Nob has to believe in order to make their respective assertions appropriately. In normal cases we infer that Hob has only to believe the content of what he asserts himself: that there is someone who killed Bob’s mare; while Nob can use the pronoun appropriately only if he also believes what is asserted by Hob. That is, Nob has to believe that there is someone who killed Bob’s mare and Cob’s cow. This asymmetry can be readily explained on standard dynamic semantics (see Groenendijk et al., 1996) if we make one extra assumption. This is that if a hearer does not respond, we can assume that he has accepted, and thus believes, what has been asserted by the earlier speaker. With the help of this assumption we can infer for this
situation of cross-speaker anaphora, i.e. a Hob-Nob situation, that Nob believes, after the update of his belief state with what is asserted by Hob, that there is someone who killed Bob’s mare.

A similar explanation can be given for the asymmetry in (20) if we assume that the pronoun is an E-type pronoun and is used as an abbreviation for the definite description recoverable from the antecedent sentence. In this case we would predict that for Nob to make his assertion appropriately, he has to believe that the one who killed Bob’s mare also killed Cob’s sow.

In the ideal case, both what is expressed by the second speaker and the asymmetry between what it is necessary for Hob and Nob to believe for each of them to make their assertions appropriately in the above discourse, can be explained straightforwardly by means of both standard dynamic semantics and the E-type approach. But standard dynamic semantics is a bit more general; it can also explain similar cases of asymmetry where no pronouns are involved. Given the very similar behavior of pronouns and presupposition triggers, we can expect the same pattern for presupposition triggers. And indeed that is what we find in cases like the following one (where [My]_F indicates that my has focal accent):

(21) Hob: My parents are gone.
       Nob: [My]_F parents are gone too.

First, when we interpret the second sentence with respect to a context updated by what Hob said, this context will satisfy the presupposition triggered by Nob’s utterance that the parents of somebody other than Nob are gone. Second, from Nob’s utterance (but not from Hob’s), we infer that Nob (but not Hob) believes that both her and someone else’s parents are gone.

5.3 Intentional identity

Notice that the asymmetry between what the first and the second speaker typically have to believe to make their respective assertions appropriate in the above cases of cross-speaker anaphora is very similar to the asymmetry between what is ascribed to Hob and to Nob, respectively, in the intentional identity attributions (8) and (9) discussed in section 3. This suggests that we can also account for the asymmetry that shows up in intentional identity attributions or Hob-Nob sentences, with either an E-type approach or a straightforward extension of standard dynamic semantics. And anyway, in intentional identity attributions we have to deal with anaphoric dependencies across sentence boundaries – a phenomenon for which the E-type approach and dynamic semantics were invented – if these sentences are to be interpreted in an incremental way.

It is relatively clear how the E-type approach would account for both (i) the possibility of anaphoric dependencies across belief attributions, and (ii) the observed asymmetry. Notice
that the E-type approach can also account straightforwardly for the variable aboutness problem of attitudes de re, as exemplified by the truth of (10) and falsity of (11).

It is also easy to imagine how dynamic semantics should be extended to account for the possibility of intentional identity attributions and their observed asymmetric behavior. We have seen above that on a dynamic semantic account every sentence (i) is interpreted with respect to a context represented by a set of world-assignment pairs; and (ii) creates a new context, the context resulting from the earlier context updated by the current sentence. Later sentences can then be interpreted with respect to this later context. The idea now is to do something similar for embedded clauses in attitude ascriptions in case of intentional identity attributions. The only difference is that embedded clauses in attitude attributions do not have to be interpreted with respect to the main context, but only with respect to a subordinated context (see the entry on modal subordination); and they create contexts with respect to which only subsequent embedded clauses, rather than entire assertions, have to be interpreted. Accordingly, embedded clauses should not be interpreted with respect to the main context, but rather with respect to subordinated contexts introduced into the discourse by the interpretation of an earlier embedded clause.

Roberts (1987) already suggested to make use of modal subordination approach to account for anaphoric and presuppositional dependencies in attitude attributions for the one-agent case, and Asher (1987), Geurts (1995, 1998) and Van Rooy (1998) worked out this suggestion.11 Of those proposals, Geurts’s analysis is the most straightforward, and we focus on that here. In his analysis, formulae representing attitude attributions are interpreted with respect to old information states and set up new ones. These old and new information states are then indexed by propositional discourse referents. Thus the intentional identity attribution like (1), repeated here as (22), can be represented by the formula in (23):12

(22) Bill believes that a woman stole his car.

He also believes that she is hiding from the police.

(23) $Bel_q(b, \exists x[Wx \land SBCx]) \land Bel_l(b, HPx)$,

Here $q$ denotes the context of interpretation with respect to which the embedded clause $\exists x[Wx \land SBCx]$ is interpreted, and $s$ the newly introduced subordinated context. The context denoted by $s$ will contain information about the variable $x$, and associates with it the information that it is a woman who stole Bill’s car. Because the second embedded clause is interpreted with respect to this newly created context, the pronoun, represented by a free variable, can be interpreted. To determine whether or not a belief attribution represented by $Bel_q(b, A)$ is true in a given world or not, Geurts assumes that the belief state of an agent is represented by a set of possible worlds, and that the above formula is true in $w$ iff for every world $v$ consistent with what $b$ believes in $w$, there is an assignment $h$ such that $⟨v, h⟩$ is an element of the context denoted by $q$ updated by $A$. The discourse (22) is predicted to be true iff the formula $Bel(b, \exists x[Wx \land SBCx \land HPx])$ is true.
Geurts uses his framework to account only for single agent cases of intentional identity attributions, but of course we might use his analysis for multi-agent cases too. Note that if we do so, we can immediately explain the asymmetry between (8) and (9), represented here by (24) and (25) respectively:

(24) $\text{Bel}_q^s(h, \exists x \text{KBM}x) \land \text{Bel}_q^s(n, \text{KCS}x)$

(25) $\text{Bel}_q^s(h, \exists x \text{KCS}x) \land \text{Bel}_q^s(n, \text{KBM}x)$

For readers familiar with discourse representation structures, (24) may be brought out more succinctly in:

(26) $\begin{array}{c|c|c|c|c}
    h, n, s, t \\
    \hline
    s: & x & t: & x \\
    & \text{KBM}x & & \text{KCS}x \\
    \hline
    \text{Bel}(h, s) \\
    \text{Bel}(n, t)
\end{array}$

In this way we predict that (i) the pronoun he in (8) satisfies the indefinite description someone who killed Bob’s mare, and (ii) that (8) is true in a situation where Nob, but not Hob, has a one-murderer theory, i.e. believes that the same person killed both Bob’s mare and Cob’s sow.

Geurts (1995) assumed that we should introduce propositional discourse markers only when embedded clauses are interpreted. But if we also treat presupposition as a propositional attitude, we have a (distinguished) propositional discourse marker, $p$, that represents what is presupposed in each possibility. What is important is that once we assume that possibilities also contain the information that is presupposed, we can also account straightforwardly for the variable aboutness problem of attitudes de re as discussed in Section 3. Remember that $p$ denotes what is presupposed in the main context. If we represent (10) and (11) by (27) and (28) respectively, we predict correctly that (10) is true and (11) false, in the situation described in Section 3 of this paper, because Hob does not have a one-murderer theory.

(27) $\exists x \text{KBM}x \land \text{Bel}_q^p(h, \neg \text{KCS}x)$

(28) $\exists x \text{KBM}x \land \text{Bel}_q^p(h, \text{KCS}x)$
Just as in the above Hob-Nob situation, also for Hob-Nob sentences the dynamic semantic solution is more general than the E-type approach. By extending dynamic semantics as above we can explain not only this asymmetry with respect to pronouns, but also when (other) presupposition triggers are involved. Consider the following example, adapted (with a change of names) from Heim (1992):

(29) a. Hob is sure that his parents are gone.

b. Tob thinks that [her]_F parents are gone, too.

In an utterance of (29b), with focus accent on her, it seems that too may relate to the information that Hob’s parents are gone and not to the information that Hob thinks that his parents are gone. On such an analysis, I think, it need not be presupposed that Hob’s parents are gone, but the sentence gives rise to the expectation that Top believes that Hob’s parents are gone. Notice that this expectation can be explained straightforwardly by means of modal subordination.

6 Problems for descriptive approaches

In the previous section we saw how both the E-type analysis of pronouns and (a straightforward extension of) standard dynamic semantics can account for the asymmetry between what Hob and Nob have to believe (i) in order for them to make appropriate assertions when they are engaged in a conversation, and (ii) to account for the fact that the belief attribution (8) is true on its most straightforward reading, while (9) is false. Indeed, it seems that the two theories make pretty good predictions. The predictions are also in accordance with Dennett’s (1968) claim that for there to be intentional identity, the agents need to descriptively characterize the putative target in the same way. But the problem is that this is the case only if certain ideal conditions hold. Unfortunately, ideal or normal conditions do not always obtain.

6.1 Cross-speaker anaphora

Consider first the case of cross-speaker anaphora, or Hob-Nob situations. Ideal conditions need not obtain here, for instance, because in the following dialogue what Nob says might well be interpretable, even if there is no man running through the park:

(30) Hob: A man is running through the park.

Nob: He wears Nike sport-shoes.

Of course, what Hob has to believe to make his assertion appropriate still has to be the same as in the ideal case; and of course, in this non-ideal situation, the first sentence, and thus the whole discourse, will not be true.
What is interesting, though, is that whether or not Nob believes or accepts what Hob says, it seems that what Nob asserts himself can be true, even if what Hob says is false. If this is indeed the case, we can conclude that the referent of a personal pronouns doesn’t have to satisfy the (in)definite description recoverable from the antecedent indefinite.

To make these cases clearer, let’s look at an example of pronominal contradiction (Strawson, 1952). Consider the following dialogue:

(31) Hob: A man is running through the park.
    Nob: He is not a man, but just a boy, and he is not running, but just walking.

Such examples differ from the ideal case in two ways: First, although Nob is saying something coherent, we cannot determine the proposition expressed by him by assuming that the pronoun satisfies the description a/the man who is running through the park, for that would give rise to the impossible proposition. Second, to be able to make this assertion appropriately, Nob also cannot believe that the ‘referent’ of the pronoun is a/the man who is running through the park.

So, although the truth value and the appropriateness of what Nob asserts are dependent somehow on Hob’s speech act, this dependence cannot be explained in the most obvious way known from the E-type approach or from dynamic semantics. In particular, we cannot interpret Nob’s assertion as the update of the initial context with Hob’s assertion without making use of some revision.

At first sight it might seem obvious how to handle cases of pronominal contradiction in standard dynamic semantics. According to these theories, we do two things when we update an initial context with what is asserted by Hob: (i) we introduce a discourse referent induced by the indefinite a man, and (ii) we associate with this discourse referent the descriptive material “being a man walking in the park”. In terms of Discourse Representation Theory (DRT), this would result in the following DR-structure:

\[
\begin{array}{c}
\text{x} \\
\text{Man(x)} \\
\text{Running – through – park(x)}
\end{array}
\]

When a second speaker uses a pronoun whose denotation depends on the indefinite used by the first speaker, but denies the descriptive material associated with it, we might say that Nob’s assertion that it is a boy who is walking in the park should be interpreted with respect to the earlier context from which the descriptive material has been eliminated.

\[
\begin{array}{c}
\text{x} \\
\vdots
\end{array}
\]
No proponent of standard dynamic semantics has ever made this proposal. The reason is that the only information that these theories associate with a discourse referent is the existential information that something exists, which is also all that is represented by the above DRS. But this information by itself will not be enough to explain the appropriateness of the dialogue in (31).

6.2 Intentional identity

Ideal conditions do not always obtain in intentional identity attributions either, as observed by Geach (1965). Geach only discussed an analysis of pronouns as abbreviations for definite descriptions recoverable from the antecedent clause, but his argument immediately carries over to the analysis of pronouns as in standard dynamic semantics. Geach argued against the descriptive approach because the second agent need not believe all of the descriptive material recoverable from the antecedent sentence. In Geach’s original sentence, for example, Nob doesn’t have to believe that the witch that he is thinking about blighted Bob’s mare, nor that Hob believes this. It seems that intentional identity attributions can be truly and appropriately made even if the agents disagree about the descriptive content associated with the belief attribution.

The Gotham city newspapers have reported that a witch, referred to as “Samantha”, has been on quite a rampage. According to the article she has been blighting farm animals and crops and throwing people down wells. In reality, there is no such person: the animals and crops all died of natural causes, and the people found at the well-bottoms had all stumbled in by accident in a drunken stupor. The news reporters simply assumed that a witch was responsible for all the mishaps, and dubbed her “Samantha”. Hob and Nob both read the Gotham Star and, like most folks, they believe the stories about the witch. Hob thinks Samantha must have blighted Bob’s mare, which took ill yesterday. Nob thinks Samantha killed his friend Cob’s sow. Nob has no beliefs at all about Hob or about Bob’s mare; he is unaware of the existence of either. (Edelberg, 1986, p. 2)

Note how similar this problem of intentional identity attributions is to the problem of pronominal contradiction which plagues classical dynamic semantics. In neither case can pronouns be treated as in the popular analyses of pronouns, as satisfying indefinite or definite descriptions recoverable from the antecedent clause. And just as in the pronominal contradiction case, the most obvious move here is to interpret the second embedded clause not with respect to the context resulting from the update of the whole of the first embedded clause, but with respect to the context resulting from the update of only part of the first embedded sentence. Instead of representing the intentional identity attribution (3) by

\[(34) \ Bel^+_q(h, \exists xWx \land BBMx) \land Bel^+_t(n, KCSx),\]


we can now represent it by

\[(35) \ Bel^x_h(\exists x Wx) \land Bel^x_h(BBMx) \land Bel^x_n(KCSx).\]

As a result, we predict that Nob does not have to believe everything that is attributed to Hob, just as we want.

But, as in the case of pronominal contradiction, this move won’t work here either. The reason is the same in both cases: according to standard dynamic semantics the only information associated with a discourse referent is existential information, which is too weak to account for the data. In this case it is too weak because it makes belief attributions too easily true. Because Nob agrees with almost nothing that is attributed to Hob, almost none of the descriptive material occurring in the embedded clause of what is attributed to Hob can occur in the description that the pronoun is going proxy for. As a result, the indefinite description will not be much richer than someone, which can hardly be enough to explain why the intentional identity attribution could be used appropriately in the first place. At this point one might suggest that the analysis can be strengthened by also demanding that their should exist a causal link between Hob’s and Nob’s attitudes. As discussed in section 2, however, such causal links are not required in general for intentional identity attributions to be appropriate. Sometimes it is just the reporter who is responsible for the link.

7 Speaker’s reference

Is it relevant to semantics whether the speaker has a certain individual ‘in mind’ by his use of the indefinite in a discourse like (36a)-(36b)? And if so, how?

(36) a. *A man* is walking in the park.

    b. *He* is whistling.

Dynamic semantics and E-type theorists say ‘no’, but according to Chastain (1975) and Donnellan (1978), among others, it is relevant both to the proposition expressed by the sentence in which the indefinite occurs, e.g. (36a), and to the propositions expressed by sentences with pronouns that take this indefinite as its syntactic antecedent, e.g. (36b). Kripke (1977), Lewis (1979), Stalnaker (1998) and Van Rooy (1997, 2001) have argued, instead, that speaker’s reference is relevant to semantics, but only through pronominalization.

It is commonly assumed (e.g. Strawson, 1952) that the phenomenon of pronominal contradiction shows that anaphoric pronouns can at least sometimes be used referentially. When Hob asserts (37a), Nob may react by saying (37b):

(37) a. Hob: *A man* jumped off the bridge.
b. Nob: He didn’t jump, he was pushed.

In these cases the pronoun appears to be used referentially, referring to the speaker’s referent of John’s use of the indefinite.

The following example,\textsuperscript{14} which illustrates what I will call the specificity problem, suggests that pronouns are more generally used referentially. If Hob utters (38a), it would be odd for him to reply to Nob’s question (38b) with (38c) if two men called Hob up yesterday and he knows this.

(38) a. Hob: A man called me up yesterday.

b. Nob: Did he have a gravelly voice?

c. Hob: That depends: if he called in the morning he did,
but if he called in the afternoon, he did not.

It not easy to see how this phenomenon can be explained if it is assumed that pronouns should simply be treated as variables bound by (dynamic) existential quantifiers. It also seems clear that the phenomenon cannot be explained by just assuming that, by Gricean Quality, speakers have to believe what they say, and account for this in terms of classical entailment; (38a) and (38c) are wrongly predicted to be acceptable given that Hob knows that two men called him up yesterday, one in the morning and one in the afternoon. A more specific relation than classical entailment between what is believed and what is said is needed to account for the intuition that Hob just wants to talk about one of the two men. A natural explanation can be given if it is assumed that for the use of the pronoun the speaker must have a specific object ‘in mind’.

In the appendix I will show how things can be done formally. Here I will just illustrate how sentences are represented and provide an informal description how they are interpreted. Consider the following discourse: A man jumped off the bridge. He died. This discourse will be represented by $JoB(\theta x(\text{Man})) \land \text{Died}(x)$, where the indefinite ‘a man’ is represented by the theta-term $\theta x(\text{Man})$ and the pronoun by $x$. The theta-term introduces a speaker’s referent to the discourse, which is picked up in the following sentence. The referent of the theta-term, the speaker’s referent, need not have jumped off the bridge. Still, according to the formal analysis, the first sentence is predicted to be true with respect to $w$ and reference context $(\phi, g)^{15}$ iff there exists a man who jumped off the bridge: $I_w(\text{Man}) \cap I_w(\text{JoB}) \neq \emptyset$.

The second sentence is true if the speaker’s referent of the indefinite, $\phi(I_w(\text{Man}))$, that is picked up by the anaphoric pronoun ‘he’ represented as $x$ died in $w$: $\phi(I_w(\text{Man})) = g(x) \in I_w(\text{Died})$, where $\phi$ is a choice function. Notice that because the speaker’s referent of the indefinite of the first sentence need not have actually jumped off the bridge, a second speaker might truly react by saying that he didn’t jump, but was pushed, which shows that we can also account for pronominal contradiction examples.

The approach sketched here can also account for the specificity problem discussed above, because it is assumed that, on its most straightforward reading, John must have a specific
object ‘in mind’ for his use of the indefinite and pronoun. When John asserts (38a), *A man called me up yesterday*, we predict that John makes the specific individual that he had in mind for his use of the indefinite *a man* available for reference for pronouns and (other) short descriptions. It would normally be odd for him to answer the question (38b), *Did he have a gravel voice?* by saying (38c), *That depends: if he called me up in the morning he did, and if he called me up in the afternoon, he did not*, because it can be assumed that the individual that the speaker had in mind as antecedent was either the one who called him up in the morning or the one who called him up in the afternoon.

At this point the obvious claim to make would be that the notion of speaker’s reference is also crucial to account for the appropriate use of many Hob-Nob sentences.16

One way to go would be to say that the indefinite antecedents used in Hob-Nob sentences will have *wide scope* with respect to the verb of belief. But as noted already in Section 2, this solution seems problematic; this is because for a sentence like (3) it does not guarantee that in all of Hob’s belief alternatives there is a witch who blighted Bob’s mare. We had seen how Saarinen (1978) proposed to solve this problem, on the assumption that indefinites should be treated as existential quantifiers, by splitting the existential quantification itself from the descriptive contribution (that of being a witch).

In this section I will not assume, however, that the ‘contribution’ of indefinites to the discourse should be split in this way. I will give the indefinites used for intentional identity attributions *narrow scope* with respect to the belief predicate. But because I will assume that when we interpret a belief attribution in world $w$ with respect to reference context $⟨φ, g⟩$ we should analyze the indefinites occurring in the embedded clause with respect to $φ$, what I propose will as far as the interpretation is concerned turn out to be roughly equivalent to the ‘*wide scope + split*’ analysis suggested above.

If we take the notion of speaker’s reference seriously, and also don’t assume that embedding under a belief blocks anaphoric binding, we no longer need modal subordination to account for intentional identity attributions. We can now represent our problematic intentional identity ascriptions simply as follows:

$$ (39) \quad Bel(h, BBM(θxW)) \land Bel(n, KCSx), $$

As a result,17 our problematic Hob-Nob sentence is predicted to be true when (i) Hob has the *existential* belief that there is a witch that blighted Bob’s mare, and (ii) that the instantiation of the *specific* (non-actual) object that the *speaker* had in mind for his use of the antecedent indefinite killed Cob’s sow in each of Nob’s belief alternatives.

Notice that the problem that arose for the earlier analysis does not arise now. The earlier analysis didn’t work because it was *too weak*, making belief attributions too easily true. Because Nob agrees with almost nothing attributed to Hob, the pronoun will satisfy almost none of the descriptive material occurring in the embedded clause attributed to Hob. As a result, it will be hard to explain the appropriate use of the intentional identity attribution. When we assume that speaker’s reference is relevant here, we can explain why
the intentional identity attribution can be used appropriately in such cases after all. The reason is that we can now associate more than just existential information with a discourse referent; the discourse referent can now be the specific (though perhaps non-actual) witch that plays in the relevant worlds the “Samantha”-role.

In the previous sections I have stressed the similarity between Hob-Nob situations and Hob-Nob sentences. But there is also an important difference. The difference is that if Hob and Nob are engaged in a conversation, it is Hob and Nob *themselves* who are responsible for their use of pronouns and presupposition triggers; but that when a speaker attributes beliefs to Hob and Nob, it is *the speaker* who is responsible for the anaphoric and presuppositional links, and not the agents that the belief attribution is about. \(^{18}\)

Our above discussion seems to have led to the following conclusions with respect to intentional identity attributions: When Hob-Nob sentences are uttered out of context, there seems to be an *asymmetry* between what the attitude subjects each have to believe in order to make the attribution true: Nob, but not Hob, has to believe what is attributed both to him and to the other. This asymmetry can be accounted for by any of the descriptive approaches towards anaphora. However, when enough context is given, the second agent doesn’t have to believe what is attributed to the first agent in order for Hob-Nob sentences to be used appropriately, and for these cases that don’t show the asymmetry it is important that pronouns can also be used referentially. Thus, asymmetric behavior should be explained by a descriptive use of pronouns, whereas non-asymmetric behavior should be explained in terms of referential pronouns.

But this cannot be the whole story. The reason is that Hob-Nob sentences also show asymmetric behavior whenever the second agent does not believe everything that is attributed to the first agent. Consider the following variant of the Hob-Nob story that pointed to the asymmetry, again due to Edelberg:

**Monday:** Smith and Jones have been shot, at opposite ends of Chicago. Arsky and Barsky are investigating both cases, but neither knows that Smith is the mayor or that Jones is the commissioner. Smith and Jones, though hospitalized, are (and are known by both detectives to be) still alive. Arsky and Barsky have discussed the two cases at length, and though they think someone shot Smith and that someone shot Jones, both believe the two cases are entirely unconnected. At this time, neither has anyone in mind as a suspect.

**Tuesday:** Both Smith and Jones have died of their gunshot wounds. Arsky knows Smith died, and thus now believes that the person who shot Smith murdered him, but doesn’t know Jones is dead. Likewise, Barsky knows Jones died, and thus now believes that the person who shot Jones murdered him, but doesn’t know Smith is dead. After reflecting on certain similarities between the two cases, Barsky infers that the man who shot Smith is the same person as the man who shot Jones. He communicates this to Arsky, saying, “The man who
shot Smith is the man who shot Jones.” Arsky disagrees, but Barsky persists in his opinion. (Edelberg, 1986, pp. 16-17)

On Tuesday, (40) is true and (41) is false on their most natural readings:

(40) Arsky believes that someone murdered Smith, and
Barsky believes he murdered Jones.

(41) Barsky believes that someone murdered Jones, and
Arsky believes he murdered Smith.

However, this asymmetry cannot be explained by treating the pronouns as abbreviations for descriptions recoverable from the clause in which the indefinite occurs. Barsky does not believe that Smith was murdered, and Arsky does not believe that Jones was murdered. The asymmetry can be explained, however, if we assume that the speaker, when he utters (40), has in mind the concept corresponding to ‘the one who shot Smith’ – something that seems plausible if both speaker and hearer are aware of the story above.\footnote{19}

8 Conclusion

Our discussion of intentional identity attributions in this paper has led to certain conclusions. First, and foremost, to analyze intentional identity attributions we need to take the notion of speaker’s reference seriously. Second, the problem of intentional identity is not just a problem of anaphora; presuppositions show the same dependencies. Third, examples of anaphora across attitude clauses are similar to examples of cross-speaker anaphora, but not the same. The difference is due to the fact that it is the speaker who is responsible for the anaphoric and presuppositional dependencies.

References


[34] Priest, G. (2005), Towards Non-being: The Logic and Metaphysics of Intentionality, Oxford University Press.


Appendix

In this appendix I will sketch how a dynamic semantics with speaker’s reference can be accounted for formally (see e.g. Van Rooy (2001) for a more elaborate discussion, including an analysis of donkey sentences). I will define a dynamic semantics for a formal language that is meant to represent a natural language, but whose syntax is close to first-order predicate logic, with the only difference that not only variables will be treated as terms, but also the added discourse markers and theta terms. I will say that $\theta x P$ is a theta term, if $P$ is a one-place predicate in which $x$ does not occur, and $x$ a variable. Indefinites will be represented by theta terms, and unbound pronouns by variable. Variables will, as usual, be interpreted with respect to partial assignment functions. To account for the intuition that speakers have specific individuals in mind for their use of indefinites that can be picked up by anaphoric pronouns, we have to guarantee that a unique individual (if there is one) is assigned to the theta term by the possibility with respect to which it is interpreted. For this reason we will assume that a possibility also contains a reference function, $\phi$, that assigns to a theta term in each world the specific individual that intuitively is the unique speaker’s reference of the indefinite (if there is one) represented by the term. Limiting ourselves to singularly used indefinites and pronouns, we can say that each reference function $\phi$ is a function from properties, i.e. functions from worlds to subsets of our domain $D$ of all possible (and perhaps non-actual) individuals, to element of $D$. Thus, terms will be interpreted as follows:

- $[[t]]^w_{\phi,g} = g(t)$, if $t$ is a variable,
- $= \phi(I(P))$, if $t = \theta x P$ and $\phi(I(P)) \in I_w(P)$,
- $= *$ otherwise

Notice that $\theta x P$ denotes $*$ in $w$ in case $\phi(I(P))$ in an object that is no $P$ in $w$.

Although, as we will see, our analysis will be truth-conditional, it will be very much dynamic in the sense that we allow some expressions to introduce discourse referents, or variables, to the discourse. To account for the referential treatment of pronouns, and the existential treatment of indefinites, we will define the notion of truth of a sentence. But to define this notion we first have to determine how theta terms can introduce variables
to the discourse and when a sentence is *rigidly true*. We will determine both separately,\(^{21}\) beginning with the definition of \(\text{Upd}(E, \langle w, \phi, g \rangle)\), which tells us how the partial assignment function \(g\) is updated after the interpretation of expression \(E\). Just to sketch the idea, I will assume that the language contains only one-place predicate constants.

\[ \begin{align*}
\text{Upd}(P(t), \langle w, \phi, g \rangle) &= \text{Upd}(t, \langle w, \phi, g \rangle) \\
\text{Upd}(t, \langle w, \phi, g \rangle) &= g, \text{ if } t \text{ is a variable,} \\
\text{Upd}(A \land B, \langle w, \phi, g \rangle) &= \text{Upd}(B, \langle w, \phi, \text{Upd}(A, \langle w, \phi, g \rangle) \rangle)
\end{align*} \]

Where \(g[x/\phi(I(P))]\) is \(g\) extended with an assignment of the speaker’s referent of the indefinite to the variable \(x\).

Now we can define the notion of *rigid truth*, where we determine the truth of the sentence with respect to a possibility in which no reference functions are existentially quantified over. I will limit myself here to the two most important clauses.

\[ \begin{align*}
[[P(t)]]^{w,\phi,g} &= 1 \text{ iff } [[t]]^{w,\phi,g} \in I_w(P) \\
[[A \land B]]^{w,\phi,g} &= 1 \text{ iff } [[A]]^{w,\phi,h} = 1 \text{ and } [[B]]^{w,\phi,h} = 1, \text{ where } h = \text{Upd}(A, \langle w, \phi, g \rangle)
\end{align*} \]

Finally, we define the notion of truth of sentence \(A\) with respect to \(\langle w, \phi, g \rangle\), \(\langle w, \phi, g \rangle \models A\), by existentially quantifying over the set, \(\Phi\), of reference functions:

\[ \langle w, \phi, g \rangle \models A \text{ iff } \exists \psi \in \Phi \text{ such that } [[A]]^{w,\psi,g} = 1. \]

Notice that when \(A\) is interpreted after the sequence of sentences \(S_1\) to \(S_n\), \(A\) will be true with respect to \(w\) and \(\phi\) if and only if there exists a man who jumped off the bridge: \(I_w(Man) \cap I_w(\text{Job}) \neq \emptyset\), and the speaker’s referent of the indefinite died: \(\phi(I(Man)) \in I_w(Died)\). Notice that because the speaker’s referent of the indefinite of the first sentence need not have actually jumped

\[ ^2 \text{ For simplicity I have assumed, and will continue to assume, that the reference function remains the same. But we might also say that the reference function should change after the interpretation of theta term } \theta xP, \text{ to account for a sequence like ‘A man called me this morning, and a man called me this afternoon.’ One way one might go (cf. Van Rooy, 1997) is to say that the new reference function, } \psi, \text{ will be exactly like } \phi, \text{ except that } \psi(I(P)) = \phi(I(P) - \{\phi(I(P))\}) \text{ to force a different choice from } I(P), \text{ where ‘−’ has the obvious interpretation.} \]
off the bridge, a second speaker might truly react by saying that he didn’t jump, but was pushed, which shows that we can also account for pronominal contradiction examples.

As for anaphora in belief contexts, we now interpret a clause of the form Bel(a, A) as given below (in the following, K(a, w) denotes the set of worlds compatible with what a believes in w).

\[\begin{align*}
&\text{• } [[\text{Bel}(t, A)]]^{w,\phi,g} = 1 \text{ iff } \forall \nu \in K([[[t]]^{w,\phi,g}, w) : [[A]]^{\nu,\phi,g} = 1 \\
&\text{• } \text{Upd}(\text{Bel}(t, A), \langle w, \phi, g \rangle) = \text{Upd}(A, \langle w, \phi, \text{Upd}(t, \langle w, \phi, g \rangle) \rangle)
\end{align*}\]

It is easy to see that the (possible) object introduced by the theta term \(\theta x P\) depends on the actual reference context, and that this possible object can be picked up by a pronoun in a later belief attribution.\(^{22}\)

We can now represent our problematic intentional identity ascriptions simply as follows:

\[\text{(42) } \text{Bel}(h, BBM(\theta x W)) \land \text{Bel}(n, KCS x),\]

As a result, our problematic Hob-Nob sentence is predicted to be true when (i) Hob has the existential belief that there is a witch that blighted Bob’s mare, and (ii) that the instantiation of the specific (non-actual) object that the speaker had in mind for his use of the antecedent indefinite killed Cob’s sow in each of Nob’s belief alternatives.\(^{23}\)
Notes

1. Thanks to a reviewer for correcting my previous statement of the problem as if it is crucially about anaphora. Two authors who clearly separated those issues are Manning (2015) and Sandgren (2016).

2. See also Dekker & Van Rooy (1998).

3. Montague himself mentions a single subject version of the problem, though attributed it to Kaplan rather than Geach.

4. Or so my mother always told me.

5. See also Zeevat (1996).

6. It is always an option to claim that semantically not only (8) but also (9) is true, and to account for the unacceptability of (9) in terms of pragmatics. I won’t discuss this proposal further, because I believe that in our cases pragmatics should help to determine what is said, and not just whether what is said is appropriate or not.

7. A similar approach towards de re belief attributions has also been worked out in van Rooy (1997) and Aloni (2001).

8. Indeed, I learned from a reviewer of this paper that something along those lines was recently proposed by Glick (2012). See also van Rooy (2000).

9. Evans (1977) claimed that the pronoun rigidly refers to (all) the $S$ such that $P$, while Neale (1990) argues that the pronoun goes proxy for the description.


11. Asher (1987) uses DRT and Van Rooy (1998) uses Van den Berg’s (1996) framework of Dynamic Plural Predicate Logic. The intuition of Van Rooy (1998) is that modal subordination has a lot in common with the dependencieς of the meanings of pronouns that ‘refer back’ to earlier indefinites in the scope of quantified phrases. Van Rooy (1998) argues that this analysis is to be preferred to other analyses of modal subordination, and he discusses also intentional identity attributions within this framework. In fact, his analysis can account for examples that King (1993) calls ‘Generalized Geach Sentences’ like ‘Winston believes that every successful movie star is a famous female actor and Emmett believes she is responsible for its success.’ Unfortunately, the framework is complicated (just as King’s (1993)).

12. I use the FCS/DPL framework rather than the DRT framework that Geurts uses. Although the choice of framework (representational or not) is important for Geurts’ analysis of presuppositions, it is not crucial for the examples that we will discuss.

13. The intentional identity example for the straightforward modal subordination account also has its presuppositional counterpart (see Dekker & Van Rooy (1998)).

14. This example came up in a discussion with Paul Dekker and Ede Zimmermann around 1996.

15. The function $\phi$ helps to select the speaker’s reference of each token of an indefinite, and the assignment function helps to determine the denotation of the pronoun.


17. See the appendix again for a more formal analysis.


19. I believe that this account is compatible with Sandgren’s (2016) ‘triangulation’-analysis. Also Sandgren notes that to explain why the beliefs of the various agents have a common focus, we sometimes need to
point to the reporter of the attitudes descriptions. Sandgren (2016) concentrates on the more abstract problem of intentional identity, and not on its linguistic manifestation, which was the focus of our paper.

\[20\text{Notice that } \theta \text{ is not a variable binder that binds occurrences of } r \text{ in } P; P \text{ doesn’t contain } x.\]

\[21\text{But in the end the two notions have to be defined simultaneously.}\]

\[22\text{In an earlier version of this paper I assumed that indefinites introduce individual concepts to the discourse. Partly due to some comments by a reviewer, I simplified the analysis allowing only for (possibly non-actual) objects to be introduced. Still, to account for single agent attributions like (1) and (2), the introduction of (possibly non-rigid) concepts might have been more natural.}\]

\[23\text{Part of this paper is taken over from chapter 3 of my dissertation (van Rooy, 1997) and from what appeared as chapter 3 of van Rooij (2006), } Attitudes and Changing Contexts, \text{ Springer, Dordrecht. I would like to thank Springer for allowing me to use this material for this paper. Furthermore, I would like to thank the reviewers of this paper and the editors of this book for many very useful comments and suggestions.}\]