Draft of March 2023.

A Relationist Theory of Intentional Identity

Dilip Ninan

dilip.ninan@tufts.edu

Abstract

This essay argues for a *relationist* treatment of intentional identity sentences like (1):

(1) Hob believes that a witch blighted Bob's mare and Nob believes that she killed Cob's sow. (Geach 1967)

According to relationism, facts of the form *a believes that* ϕ *and b believes that* ψ are not in general reducible to facts of the form *c believes that* χ. We first argue that extant, non-relationist treatments of intentional identity are unsatisfactory, and then go on to motivate and explore a relationist alternative in some detail. We show that the general thesis of relationism can be directly motivated via cases already discussed in the literature, and then develop a particular version of relationism couched in the possible worlds framework. The resulting theory avoids the problems facing its non-relationist rivals, and yields a natural account of the truth-conditions of (1). And these truth-conditions can be generated in a compositional manner by a precise version of dynamic semantics. The theory also helps us to cleanly separate semantic questions about intentional identity from metasemantic ones.

1 Introduction

Suppose that I believe that ϕ and you believe that ψ, for some ϕ and some ψ. You and I thus stand in a certain two-place relation, which we may visualize as follows:

____ believes that ϕ and ____ believes that ψ.

Let's call such a relation a *dyadic belief relation*. Suppose now that I also believe that χ, for some χ. Then I have a certain monadic property, which we may visualize as follows:

____ believes that χ.

Let's call such a property a *monadic belief property*. It is natural to think that the facts about which individuals stand in which dyadic belief relations is determined by the facts about which individuals have which monadic belief properties. Indeed, one might think that the logic of conjunction alone guarantees this. Nevertheless, I think that there is a case to be made that this natural view is wrong, and that some dyadic belief facts are in fact not reducible to the monadic belief facts. Let us call this broad view *relationism about belief*.

My argument for relationism involves the phenomenon of *intentional identity*, which Geach (1967) first introduced with examples like this:

(1) Hob believes that a witch blighted Bob's mare and Nob believes that she killed Cob's sow.

While sentence (1) may have more than one reading, Geach was interested in a reading of it on which it does not entail the existence of witches, and on which the pronoun *she* occurring in the second conjunct is in some sense anaphoric on the indefinite description *a witch* occurring in the first conjunct.

What is the connection between relationism and intentional identity? The idea is this: as the extant literature reveals, it turns out to be surprisingly difficult to find a pair of monadic beliefs, b and b', such that (1) is true iff Hob has b and Nob has b'. A tempting conclusion to draw is that the reason we can't find a pair of monadic beliefs meeting this description is that there is no such pair; in other words, the dyadic belief fact asserted by (1) is not reducible to any conjunction of monadic belief facts. But whether we ought to embrace this relationist conclusion depends on two things: (i) how difficult it really is to find the needed pair of monadic beliefs, and (ii) whether switching from non-relationism to relationism helps matters at all. While I shall present arguments that cast doubt on non-relationist approaches, my principal aim in this essay is to demonstrate that switching from non-relationism to relationism really does advance our understanding of intentional identity.

I begin in Section 2 by examining extant, non-relationist approaches to intentional identity. My aim here is not to refute non-relationism definitively, but only to impress upon the reader the difficulty of constructing a satisfactory non-relationist theory of these matters. Section 3 is then devoted to motivating and developing our relationist alternative in detail. We begin in Section 3.1 by offering a more precise statement of relationism and then providing some initial motivation for that view. In Section 3.2, we present a particular version of relationism, one couched in possible worlds semantics. We show that this approach yields a natural account of the truth-conditions of intentional identity sentences, one that avoids the problems facing non-relationist accounts. Section 3.3 shows how our proposal helps to separate semantic questions about intentional identity from metasemantic ones, and Section 3.4 develops a version of dynamic semantics that assigns to intentional identity sentences our proposed relationist truth-conditions.

2 Non-relationism

2.1 Descriptivism

To see the sort of reading of (1) that Geach has in mind, imagine Hob and Nob having the following sort of exchange:

> HOB: There's a witch going around town these days. I think she blighted Bob's mare last night.
>
> NOB: I heard about that witch. I bet she also killed Cob's sow.

Focussing on such cases suggests that Geach's sentence is equivalent to something like the following:

(2) Hob believes that a witch blighted Bob's mare and Nob believes that the witch that blighted Bob's mare killed Cob's sow.

The intended reading is one on which both belief ascriptions are read *de dicto*. On this approach, the pronoun *she* in (1) somehow goes proxy for the underlined definite description in (2). Alternatively, perhaps Nob is unsure as to whether the witch in question really did blight Bob's mare, but nevertheless realizes that *Hob believes* that she did. In that case, we might instead interpret Geach's sentence as follows:

(3) Hob believes a witch has blighted Bob's mare and Nob believes that the witch that Hob believes blighted Bob's mare killed Cob's sow.

Here the pronoun in (1) is understood to go proxy for the more elaborate description *the witch that Hob believes blighted Bob's mare.* Approaches along this line fit naturally with what are known in the semantics literature as E-type approaches to anaphora.[1]

Notice that, on either of these views, (1) becomes straightforwardly equivalent to a conjunction of *de dicto* belief reports. On either view, there is is a pair of beliefs b and b' such that (1) is true iff Hob has b and Nob has b'. For instance, on the first of these views, b is the belief that a witch blighted Bob's mare, while b' is the belief that the witch that blighted Bob's mare killed Cob's sow. So it seems that, on this view, there is a clear sense in which the dyadic belief fact reported by (1) is reducible to a pair of monadic belief facts.

Now while (2) and (3) might both be possible readings of (1), Geach observed in his original article that these are not the only available readings of (1). For it seems that there are situations in which (1) is true while both (2) and (3) are false. Here is an example, adapted from Edelberg (1986):

> **Newspaper Case**
>
> A number of animals in Gotham Village have recently died quite unexpectedly. Rumors have begun to circulate that these unfortunate

[1]See, for example, Evans (1977), Cooper (1979), Heim (1990), and Elbourne (2005).

events are due to the machinations of a witch. The local newspaper, the Gotham Star, has picked up on these rumors and dubbed the witch *Samanatha*. The paper has reported that Samantha has been attacking animals and destroying crops. In reality, there is no such individual: the animals in question all died of natural causes, the crops withered from drought. Hob and Nob both read the Gotham Star and both believe the articles about the witch. Hob thinks that the witch must have blighted Bob's mare, which fell ill recently, while Nob thinks that the witch killed Cob's sow. But Nob is unaware of Hob's and Bob's existence, and so has no beliefs about Hob or Bob at all.

Since Nob has no beliefs about either Hob or Bob, Nob does not believe that the witch who blighted Bob's mare killed Cob's sow, nor does he believe that the witch that Hob believes blighted Bob's mare killed Cob's sow. Thus, neither (2) nor (3) is true in this scenario. But it is widely thought that (1) has a reading on which it is true in this scenario.[2] If that is correct, then neither (2) nor (3) captures the intended reading of (1).

A number of authors take cases like the Newspaper Case to show that the truth of (1) requires that Hob's belief and Nob's belief have a common causal source (Glick, 2012; Cumming, 2014; Lanier, 2014).[3] For note that, in the Newspaper Case, Hob's belief and Nob's belief are both partially caused by the articles in the Gotham Star, or by the rumors circulating in town. This observation can be used to motivate an alternative 'descriptivist' story.

On the alternative approach, (1) is instead equivalent to something like:

(4) Hob believes that a witch blighted Bob's mare and Nob believes that the witch described by the actual common causal source of Hob's belief and Nob's belief killed Cob's sow. (Lanier, 2014, 298)

The inclusion of the adjective *actual* is intended to make (4) (and thus (1)) equivalent to:

(5) Hob believes that a witch blighted Bob's mare and the common causal source S of Hob's belief and Nob's belief is such that Nob believes that the witch described by S killed Cob's sow.

Note that if (5) is true, then *the common causal source of Hob's belief and Nob's belief* is a non-empty definite description. Since (1) is equivalent to (5) on this proposal, the account predicts that the truth of (1) entails that Hob's belief and Nob's belief have a common causal source. Furthermore, this proposal avoids the problem facing the previous descriptivist approaches, since its proposed truth-condition doesn't require that Nob is aware of Hob or Bob, only that he is aware of the causal source of his belief.

[2] Though see King (1993) and Braun (2012) for some doubts about this.

[3] Here and in what follows, by *Hob's belief* I mean Hob's belief that a witch blighted Bob's mare, and by *Nob's belief* I mean Nob's belief that a witch killed Cob's sow.

The principal difficulty with this approach is that while (1) might require that Hob's belief and Nob's belief *have* a common causal source, it would not appear to require that *Nob believe* anything about this source. For example, while it is natural to assume that, in the Newspaper Case, Nob has a *de dicto* belief to the effect that the witch described in the Gotham Star article killed Cob's sow, a nearby variant of the case lacks this feature. Imagine, for example, that Nob reads the article in the Gotham Star, forms the belief that the witch killed Cob's sow, and then proceeds to forget how he formed this belief. Maybe he later comes to believe that he learned about the witch from his friend Janice, or maybe he simply forms no new beliefs about the source of his witch-beliefs. After all, we often forget how we formed certain beliefs, but retain those beliefs nevertheless. The article in the Gotham Star is the causal source S of Hob's belief and of Nob's belief, but since Nob has forgotten all about that article, he has no beliefs about S. So (5) is false is this version of the Newspaper Case. Nevertheless, it seems that (1) is still true, which suggests that (1) is not equivalent to (5) after all.[4]

2.2 Mythical objects

The foregoing considerations provide modest evidence in favor of relationism. For we've been struggling to find a pair of monadic beliefs b and b' such that (1) is true iff Hob has b and Nob has b', and one explanation of this fact is that there is no such pair, just as relationism would predict. But one might, of course, draw an alternative conclusion from our inability to find the needed pair of monadic beliefs. For what we've seen so far is that, if we restrict our search to beliefs concerning ordinary objects and their properties, it is difficult the needed pair of monadic beliefs. But perhaps this just shows that the class of monadic beliefs is larger than we initially thought: in addition to including beliefs whose content may be characterized by ordinary objects and their properties, it includes beliefs concerning certain kinds of *non-standard* objects and their properties.

One family of approaches to intentional identity draws precisely this conclusion. Salmon (2002), for example, holds that (1) is true iff there is a 'mythical witch' x such that Hob believes that x blighted Bob's mare and such that Nob believes that x killed Cob's sow. On Salmon's view, a *myth* is any false theory that has been held true; and a *mythical object* is a hypothetical entity erroneously postulated by a theory (Salmon, 1998, 304). Mythical objects are abstract objects; they are neither physical objects nor mental entities. Salmon's view is that whenever someone believes that there is an F that is such-and-such when there is no F that is such-and-such, then there is a mythical F thereby believed to be such-and-such.

Since Hob believes there is a witch who blighted Bob's mare even though there is no such witch, it follows that there is a mythical witch x that Hob

[4]The considerations in this paragraph also cast doubt on the counterpart-theoretic account of intentional identity found in Glick (2012). See, in particular, Glick (2012, 391) where Glick's characterization of the relevant counterpart relation relies on there being a newspaper article S such that Hob and Nob both believe that they have read S.

believes blighted Bob's mare. Furthermore, Salmon holds that if two believers believe there is an F that is such-and-such when there is no F that is such-and-such, "they may or may not believe in the same mythical F, depending on their interconnections" (Salmon, 2002, 105, n. 25). Thus, we may assume that Hob's and Nob's interconnections in the Newspaper Case are such that they believe in the same mythical witch. In that case, if we assume that (1), on the relevant reading, has the same truth-conditions as (6), then we predict that (1) is true in the Newspaper Case.

(6) There is a mythical witch such that Hob believes that she blighted Bob's mare, and Nob believes that she killed Cob's sow.

And since Hob's and Nob's interconnections in the Newspaper Case are preserved even in the variant in which we stipulate that Nob forgets how he acquired his belief, this approach likewise predicts a true reading of (1) in that variant of the case. So the 'mythical objects' approach avoids the problems facing the descriptivist views discussed above.

Of course, some philosophers will find it hard to believe that there are any mythical objects of the sort Salmon posits. If such philosophers also accept that (1) is true in scenarios like the Newspaper Case, they will have reason to reject Salmon's approach. But even if one is willing to grant that there are mythical objects, it is a separate question as to whether sentences like (1) *entail* that there are mythical objects.[5] I believe that there are chairs, but I don't believe that (1) entails that there are chairs.

To see the issue, compare the following two dialogues:

(7) *A*: There is a mythical witch such that Hob believes that she blighted Bob's mare and Nob believes that she killed Cob's sow.

B: Wow, I didn't know you believed in mythical witches.

B: Well, I don't believe that because I don't believe that there are mythical witches.

(8) *A*: Hob believes that a witch blighted Bob's mare, and Nob believes that she killed Cob's sow.

B: ? Wow, I didn't know you believed in mythical witches

B: ? Well, I don't believe that because I don't believe that there are mythical witches.

B's responses to *A* in (7) seem reasonable, suggesting that *A*'s utterance in that dialogue does indeed entail that there are mythical witches. In contrast, *B*'s response to *A* in (8) is odd; *A* might well respond by denying that she believes in mythical witches, and it seems that she would be well within her rights to so respond. This suggests that *A*'s utterance in (8)—which is simply an utterance of sentence (1)—does not entail that there are mythical witches. In fact, even

[5]Braun (2012) raises a similar objection.

Salmon himself seems to concede the point (Salmon, 2002, 107, n. 28); he just doesn't think a better analysis is available. But, as we shall see, the relationist alternative developed below avoids this consequence.

3 Relationism

3.1 The general thesis and some initial motivation

The foregoing discussion of non-relationist approaches to intentional identity has not been exhaustive, but it does suggest that finding a satisfactory non-relationist account is no simple matter.[6] That provides at least some motivation for considering relationist alternatives, and it is to this task that we now turn.

We begin by giving more precise characterizations of relationism and non-relationism, respectively. Let's say that an agent a *has precisely the same monadic beliefs* in w as they have in w' iff: for all ϕ, a believes that ϕ in w iff a believes that ϕ in w'.[7] And let's say that agents a and b *stand in precisely the same dyadic belief relations* in w as they do in w' iff: for any ϕ and ψ, (a believes that ϕ and b believes that ψ in w) iff (a believes that ϕ and b believes that ψ in w'). Then *non-relationism about dyadic belief* is the following view:

> NON-RELATIONISM: For any worlds w and w' and agents a and b, if a has precisely the same monadic beliefs in w as they have in w', and b has precisely the same monadic beliefs in w as they have in w', then a and b stand in precisely the same dyadic belief relations in w as they do in w'.

And *relationism about dyadic belief* is simply the rejection of non-relationism:

> RELATIONISM: The negation of non-relationism.

The bulk of the rest of the essay develops in detail a particular version of relationism, one couched in possible worlds semantics. But it is worth separating that particular version of relationism from the general thesis just stated, and worth observing that the general thesis can be motivated independently of the arguments for our favored version of relationism.

To see this, note that non-relationism is a particular way of saying that the dyadic belief facts supervene on the monadic belief facts. Thus, we can attempt to construct a counterexaple to non-relationism by providing a pair of

[6] One notable omission in our discussion of extant accounts of intentional identity is the approach due to Edelberg (1986, 1992, 1995) and further developed by Cumming (2014). Because this style of approach invokes a non-standard semantic appraratus (e.g. indefinite descriptions are referential, rather than quantificational, and they denote 'thought-objects' rather than ordinary objects), it would take us too far afield to examine it in any detail. Furthermore, the extant literature has already turned up some problems for this approach: Cumming (2014), for example, points out flaws in the proposals of Edelberg (1986) and Edelberg (1992), while Lanier (2013, Ch.3) raises problems for Cumming's own proposal.

[7] Here and in what follows, we employ a metalanguage that permits quantification into sentence position.

cases which differ with respect to the dyadic belief facts but do not differ with respect to the relevant monadic belief facts. Fortunately for us, we do not need to construct such a pair *ex nihilo*, since the extant literature already suggests a pair of cases which has precisely this feature.

Consider, for example, the following pair of cases, lightly adapted from Lanier (2014, 292):

> **The Connected Case**
>
> Al and Bud both suspect that a witch has come to town and is poisoning livestock and destroying crops. Al and Bud get together, discuss their respective theories, and decide to warn the town, each from a separate location. Each man goes to his designated location and begins to warn passersby: "There's a witch in town! She's poisoning our livestock and destroying our crops! Be on your guard!" Hob hears Al, and concludes that the witch in question must have blighted Bob's mare, which fell ill recently. Nob hears Bud, and concludes that the witch in question must have killed Cob's sow, which died unexpectedly last night. Of course, no witch (or any other person) caused any of the mishaps in question, all of which were due to natural causes. We may also assume that, as in the Newspaper Case, Nob knows nothing of Hob or Bob.
>
> **The Unconnected Case**
>
> This is exactly like the Connected Case, except that Al and Bud have never met in their life and have no coordinated plan to warn the townspeople about the witch. Al is delusional. Bud is bored and decides to start a witch-hunt by spreading a rumor to the effect that there is a witch in town causing trouble. Al goes to the same location that he goes to in the Connected Case, and similarly for Bud, and each man makes the same speech that he made in the Connected Case. And, again, Hob overhears Al and comes to believe that the witch Al is talking about must have blighted Bob's mare, while Nob overhears Bud and comes to believe that the witch Bud is talking about killed Cob's sow.

The Unconnected Case is essentially exactly like the Connected Case in all relevant respects, except that Hob's belief and Nob's belief do not derive from a common causal source. Lanier observes that (1) appears to be true in the Connected Case, for that case is essentially like the Newspaper Case. But Lanier also claims that (1) is false in the Unconnected Case, and Glick (2012, 392) reports a similar judgment.

Lanier's purpose in discussing pairs of cases like this is to argue for the common cause requirement. We shall return to that issue below, but here we observe that such pairs of cases can also be used to argue for relationism. Since (1) is true in the Connected Case, but not in Unconnected Case, there is a dyadic belief fact that holds in the Connected Case, but not in the Unconnected Case.

But given how the cases are described, it is natural to suppose that Hob has precisely the same monadic beliefs in the Connected Case as in the Unconnected Case, and similarly for Nob. There is, at any rate, nothing in the description of these cases that stands in the way of our simply stipulating that Hob has precisely the same monadic beliefs in both cases, and similarly for Nob. For the only relevant difference between the two cases concerns whether or not Al and Bud are colluding, a difference that need not make any difference to either Hob's monadic beliefs or to Nob's monadic beliefs. If we accept that these two cases may be filled out in this way, then Hob and Nob have precisely the same monadic beliefs in them, despite the fact they do not stand in precisely the same dyadic belief relations in them. In that case, relationism will be true, non-relationism false.

3.2 Relationism vs. non-relationism

The foregoing argument provides some initial motivation for relationism, but we can extend the case by developing relationism in more detail, and then comparing the resulting theory to the non-relationist theories discussed in Section 2. The particular version of relationism I want to propose is couched in possible worlds semantics, and so it will be useful to consider briefly how the problems we discussed in Section 2 arise within that framework.

On standard possible worlds theories of attitudes (Hintikka, 1962), for any agent a and any ϕ, a believes that ϕ iff in every world w compatible with what a believes, ϕ. Framed in that way, the initial problem posed by (1) becomes the problem of saying what it is for a world w to be compatible with what Hob believes and what it is for a world w' to be compatible with what Nob believes, given that (1) is true. The first of these questions has a natural answer: w should contain a witch who blighted Bob's mare. The trouble comes with saying what w', an arbitrary world compatible with what Nob believes, should be like. World w' should contain someone y who killed Cob's sow, but who in w' is y? We can think of the various proposals considered in §2 as offering different answers to this question. For example, the first version of descriptivism we considered says that y is the witch who blighted Bob's mare in w'. The advocate of the mythical objects view says instead that y is identical to x, and x/y is a mythical object. But, as we argued above, none of these answers is wholly satisfactory.

Now the fact that this question has proved so difficult to answer suggests that there might be something wrong with the question itself. While it is hard to be sure of this, I think we have said enough at this point to motivate considering an approach that focusses on a different question altogether. I propose that instead of asking about what it is for a world w to be compatible with what Hob believes and what it is for a world w' to be compatible with what Nob believes, we instead ask what it is for a *pair* of worlds (w, w') to be compatible with what the pair (Hob, Nob) believe. And *this* question turns out to have a comparatively natural answer: (w, w') should be compatible with what (Hob, Nob) believe only if *there is an x such that x is a witch in w, x blighted Bob's mare in w, and x killed Cob's sow in w'*. (Here, the first element of (w, w')

is indexed to Hob, the second to Nob.) The view taken here is (to a first approximation) that (1) is true iff every pair (w, w') compatible with what (Hob, Nob) believe meets the italicized condition above. This basic idea will be further developed in the remainder of the essay.

Let us start with a simple question about what we've said far: what is it for a pair of worlds (w, w') to be compatible with what a pair of individuals (a, b) believe? We can approach this question by examining the parallel question that arises for the standard possible worlds semantics for attitude reports. As we said above, the standard view holds that an a believes that ϕ iff in every world w compatible with what a believes, ϕ. How do we understand the notion figuring on the right-hand side of this biconditional, the notion of a world's being compatible with what an agent believes? The basic idea is that w is compatible with what an agent a believes iff: for all ϕ, if a believes that ϕ, then ϕ in w. So if, for example, Sam believes that it is raining in Tokyo, then w will be compatible with what Sam believes only if it is raining in Tokyo in w.[8]

The relationist can say something similar about the notion of a pair of worlds being compatible with what a pair of agents believe. The basic idea is that (w, w') will be compatible with what a pair of agents (a, b) believe iff: for all ϕ and ψ, if a believes that ϕ and b believes that ψ, then ϕ in w and ψ in w'. So if, for example, Sam believes that it is snowing in Chicago and Tomoko believes that it is raining in Seattle, then (w, w') will be compatible with what (Sam, Tomoko) believe only if it is snowing in Chicago in w and it is raining in Seattle in w'. More interestingly, if Sam believes that a senator from New England embezzled funds and Tomoko believes that she lied to the FBI, then (w, w') will be compatible with what (Sam, Tomoko) believe only if there is an x such that x is a senator from New England in w, x was embezzled funds in w, and x lied to the FBI in w'.

Now these remarks, I believe, suffice to show that the relationist's key notion—that of a pair of worlds being compatible with what a pair of agents believe—is intelligible, or at least as intelligible as the parallel notion typically taken for granted in standard possible worlds theories of attitudes. So we shall take this notion for granted in what follows, and see where doing so leads us.

Our proposal, recall, is that sentence (1) is true iff for all pairs of worlds (w, w') compatible with what (Hob, Nob) believe, there is an x such that x is a witch in w, x blighted Bob's mare in w, and x killed Cob's sow in w'. The first thing to observe about this approach is that it avoids the various problems facing the non-relationist theories discussed earlier. Note, for example, that, on this account, (1) does not entail (2):

(2) Hob believes that a witch blighted Bob's mare and Nob believes that the witch that blighted Bob's mare killed Cob's sow.

On the present approach, (2) would be true iff for all pairs of worlds (w, w') compatible with what (Hob, Nob) believe, there is an x such that x is a witch in

[8]It should be clear that nothing we've said amounts to a non-circular analysis of the notion of belief, nor is such intended. Such an analysis, if there is one, is left to the philosopher of mind or to the cognitive scientist.

w, x blighted Bob's mare in w, x is the unique witch that blighted Bob's mare in w', and x killed Cob's sow in w'. But given natural assumptions about what the space of worlds is like, this is clearly a stronger condition than the truth-condition we proposed for (1). For any pair of worlds (w, w') and individual x that witnesses the truth-condition for (1), x must blight Bob's mare in w and kill Cob's sow in w'. But that appears to be consistent with x not blighting Bob's mare in w', since w' may well be a distinct world from w.

Similarly, the present approach avoids the problem facing the 'causal descriptivist' discussed earlier. For it is clear our proposed truth-condition for (1) does not require that Nob believe anything about the source of his belief. Note also that the present approach avoids the problems facing the 'mythical objects' view, for it does not imply that (1) entails that there are mythical objects. So in these respects, our relationist approach appears to be a genuine improvement over its non-relationist rivals.

I have been calling the present account *relationist*, but I have to justify my doing so. We need to see why this account counts as relationist in the sense of Section 3.1. We argue as follows. Let a and b be fixed but arbitrary agents. We assume that we can extract the monadic belief facts from the dyadic ones in the following way. Let $Dox^w_{a,b}$ be the set of pairs of worlds compatible with what (a, b) believe at in w, let Dox^w_a be the set of worlds compatible with what a believes at w, and let Dox^w_b be the set of worlds compatible with what b believes at w. Then we assume that:

$$Dox^w_a = \{v : (v, v') \in Dox^w_{a,b} \text{ for some } v'\}, \text{ and}$$

$$Dox^w_b = \{v' : (v, v') \in Dox^w_{a,b} \text{ for some } v\}.$$

And we assume that for any $c \in \{a, b\}$, c has precisely the same monadic beliefs in w as they have in w' iff $Dox^w_c = Dox^{w'}_c$. We also assume that a and b stand in precisely the same dyadic belief relations in w as they do in w' iff $Dox^w_{a,b} = Dox^{w'}_{a,b}$. Given natural assumptions about the space of worlds, we can show that there are worlds w and w' such that $Dox^w_a = Dox^{w'}_a$, $Dox^w_b = Dox^{w'}_b$, but $Dox^w_{a,b} \neq Dox^{w'}_{a,b}$. In that case, a and b will each have precisely the same monadic beliefs in w as they have in w', but they will not stand in precisely the same dyadic belief relations in w as they do in w'. Relationism will be true, non-relationism false.

To see how this would work, suppose there are worlds w and w' such that:

$$Dox^w_{a,b} = \{(v, v') : \exists x\ (x \text{ blighted Bob's mare in } v \text{ and } x \text{ killed Cob's sow in } v')\}.$$

$$Dox^{w'}_{a,b} = \{(v, v') : \exists x\ (x \text{ blighted Bob's mare in } v) \text{ and } \exists y\ (y \text{ killed Cob's sow in } v')\}.$$

It seems that $Dox^w_{a,b} \neq Dox^{w'}_{a,b}$. For suppose that in v, x alone blighted Bob's mare, and that in v', y alone killed Cob's sow, where $y \neq x$. Then (v, v') will be in $Dox^{w'}_{a,b}$, but not in $Dox^w_{a,b}$ Thus, a and b will not stand in precisely the same dyadic belief relations in w as they do in w'. But we also have that

$Dox_a^w = Dox_a^{w'}$ and $Dox_b^w = Dox_b^{w'}$, which means that a and b each have precisely the same monadic beliefs in w as they have in w'.[9] Thus, we have a difference in the dyadic belief facts despite no difference in the relevant monadic belief facts.

3.3 Metasemantics

One thing that keeps popping up in our discussion is the idea that that the truth of an intentional identity sentence imposes a common cause requirement. This was one of the main motivations for the causal descriptivist proposal discussed in Section 2. And Glick (2012, 392) takes pairs of cases like the Connected Case and the Unconnected Case to show that (1) is true iff three conditions obtain: (i) Hob believes that a witch blighted Bob's mare, (ii) Nob believes that someone killed Cob's sow, and (iii) Hob's belief and Nob's belief have a common causal source. Suppose, for the moment, that Glick's claim is true. Then while the present proposal entails that (1) is true only if conditions (i) and (ii) hold, it does not obviously imply (iii). So, again assuming (1) does impose a common cause requirement, this raises the question of where condition (iii) fits into our analysis. I will first answer this question on the assumption that intentional identity sentences really do impose a common cause requirement, and then return to consider whether this assumption in fact holds.

My view is that the common cause requirement is most naturally understood as a *metasemantic* requirement, something that must obtain in order for Hob and Nob to stand in the dyadic belief relation that (1) says that they stand in. It is not, as the causal descriptivist maintains, something that that figures in the *content* of the psychological states reported. Compare (1) with (9):

(9) Kripke believes that Feynman was a physicist.

It may be that (9) is true only if Kripke is causally related to Feynman in an appropriate manner. But this fact does not figure in the content of the belief reported, which simply concerns Feynman and one of his properties. Instead, the causal requirement is arguably a metasemantic requirement, a requirement on what must be true of Kripke in order for him to have the property of believing that Feynman was a physicist. Similarly, if (1) does impose a common cause requirement, I suggest this fact does not figure in the content of the state that (1) attributes to Hob and Nob, but is a requirement on what must be true of Hob and Nob in order for them to stand in the dyadic belief relation that (1) reports

[9] To see that $Dox_a^w \subseteq Dox_a^{w'}$, suppose $v \in Dox_a^w$. Then there is a v' and an x such that x blighted Bob's mare in v and x killed Cob's sow in v'. But then there is there is a z that blighted Bob's mare in v and there is a y that killed Cob's sow in v', for x is such a z and such a y. So $(v, v') \in Dox_{a,b}^{w'}$, which means $v \in Dox_a^{w'}$. To see that $Dox_a^{w'} \subseteq Dox_a^w$, suppose $v \in Dox_a^{w'}$. So there is a v' such that there is an x that blighted Bob's mare in v and such that there is a y that killed Cob's sow in v'. Let u be a world in which x killed Cob's sow (we assume, plausibly, that there is such a world). So x blighted Bob's mare in v and x killed Cob's sow in u. So $(v, u) \in Dox_{a,b}^w$, which means $v \in Dox_a^w$, as desired. The argument that $Dox_b^w = Dox_b^{w'}$ is similar.

them as standing in. From the present perspective, the causal descriptivist mislocates the common cause requirement, putting it into the semantics when it is properly understood as a feature of the metasemantics.

All this is assuming that (1) really does impose a common cause requirement. Is that true? And even if that is true of sentence (1) in particular, is it generally true that intentional identity sentences impose common cause requirements? Sarah Moss (p.c.) has suggested to me that this last question should be answered in the negative. For example, imagine that we have two causally disconnected cultures, that, perhaps by chance, have very similar views theological beliefs concerning matters like the creation of the universe and the origins of humanity. If we fill in the details in the right way, this might suffice for the truth of (10):

(10) Culture A believes that a supreme being formed humans out of clay, while Culture B believes that he formed them out of fire and water.

And this despite the fact that Culture A's belief does not have the same causal source as Culture B's. If this possible, then it will not generally be true that intentional identity sentences require for their truth that the corresponding monadic beliefs derive from a common causal source. Perhaps what is driving our judgment that (10) is true in this scenario is the fact that the deity hypothesized by Culture A plays a similar explanatory role to the deity posited by Culture B.[10]

As far as I can see, the relationist *qua* relationist needn't take any particular stand on this issue. From the relationist point of view, the questions raised by this example concern the metasemantic requirements that must be met in order for a pair of subjects to stand in a particular dyadic belief relation. What the relationist has offered is an abstract account of particular dyadic belief relations themselves, not an account of the conditions that must obtain in order for such a relation to be instantiated. The relationist tells us that (1) is true iff Hob and Nob stand in the relation that a bears to b iff every (w, w') compatible with what (a, b) believe that there is a witch x who blighted Bob's mare in w and who killed Cob's sow in w'. It does not tell us what has to be true of Hob and Nob in order for every (w, w') compatible with what (Hob, Nob) believe to satisfy this condition.

Again, it is instructive to compare the present situation with the case of the standard possible worlds semantics for attitude ascriptions. On the standard account, (9), for example, is true iff every world w compatible with what Kripke believes is such that Feynman is a physicist in w. But the standard account does not, by itself, tell us what has to be true of Kripke in order for every world w compatible with what he believes to satisfy this condition.

All that being said, the present discussion might help us to see what information is being encoded by dyadic belief ascriptions, and this might cast light on why we would use dyadic belief ascriptions instead of simply relying on conjunctions of monadic belief ascriptions. For if we generalize a bit from

[10]See Edelberg (1992, §8) on the relevance of explanatory roles to the truth of intentional identity sentences.

Glick's proposal, we might suppose that (1) is true iff (i) Hob believes that a witch blighted Bob's mare, (ii) Nob believes that someone killed Cob's sow, and (iii) [INSERT FAVORED METASEMANTIC REQUIREMENT HERE]. In that case, intentional identity sentences can be seen as encoding two types of information: they tell us about the relevant subjects' respective monadic beliefs, and they also tell us that the relevant metasemantic requirement—whatever, precisely, it is—has been met. An account of this general form can be given regardless of what precise form the metasemantic requirement takes.

Note that if the preceding (schematic) view is correct, then it suggests that dyadic belief facts *can* be reduced to the corresponding monadic belief facts together with certain other facts, namely whatever other facts constitute the metasemantic requirement. Is this a source of embarrassment for the relationist, who maintains that the dyadic belief facts are not reducible to the monadic belief facts? It is not. For it is no part of the relationist's view that dyadic beliefs are irreducible *tout court*—the relationist, *qua* relationist, need not hold that dyadic beliefs are part of the fundamental furniture of the universe. Even if relationism is true, it may also be true that dyadic belief facts are reducible in some way to something, and if that is so, they are likely to be reducible to the monadic belief facts together with certain additional facts. It might be helpful to compare the situation here to the case of externalism about psychological states, in the style of Putnam (1973) and Burge (1979). The externalist denies that the monadic belief facts are reducible to certain local functional and physical facts. But that thesis is compatible with the claim that the monadic belief facts are reducible to those local facts *together with* facts about the physical and social environment (Stalnaker, 1984, Ch. 1). The externalist, *qua* externalist, needn't maintain that monadic beliefs are part of the fundamental furniture of the universe. And the same goes, *mutatis mutandis*, for the relationist.[11]

3.4 Compositional semantics

Our final task is to construct a compositional semantics that predicts our proposed truth-condition for (1). But before we get to that, we first need to generalize our account. For note that there is nothing special about *dyadic* beliefs in particular:

(11) Hob believes that a witch blighted Bob's mare, Nob believes that she killed Cob's sow, and Joe believes that she stole Janice's tractor.

That suggests that we should speak of an n-ary sequence of worlds $(w_1, ..., w_n)$'s being compatible with what an n-ary sequence of individuals $(a_1, ..., a_n)$ believe. But since nothing in the present phenomenon mandates the order built into these sequences, I propose that to use functions from agents to worlds instead of sequences.

Where A is the set of agents and W the set of worlds, let an *indexing function* be a function w mapping agents in A to worlds in W. Note that the range of any

[11]Thanks to Andy Egan and Arc Kocurerk for discussion on this issue.

given indexing function w is an indexed set of worlds $\{\mathsf{w}_{\text{Hob}}, \mathsf{w}_{\text{Nob}}, ...\}$, and it can be useful to visualize an indexing function w by picturing the corresponding indexed set.[12] Assuming $\{\text{Hob, Nob}\} \subseteq A$, I propose that (1) is true at a world w iff for all indexing functions $\mathsf{w} \in W^A$ compatible with what the agents in A believe in w, there is an x such that x is a witch who blighted Bob's mare in w_{Hob} and x killed Cob's sow in w_{Nob}.[13]

Our task now to is to provide a theory that assigns this truth-condition to (1) in a compositional manner. But this task is not completely straightforward. To get a sense of the challenges here, note first that we may regiment (1) as follows:

(R1) $\mathcal{B}_b \exists x Fx \wedge \mathcal{B}_c Gx$

Here $\mathcal{B}_b$ translates *Hob believes that*, $\mathcal{B}_c$ translates *Nob believes that*, F translates *is a witch who blighted Bob's mare*, and G translates *killed Cob's sow*. But note that our proposed truth-conditions essentially have the following form:

(T1) $\forall \mathsf{w} \in \text{Dox}_A : \exists x(Fx \text{ in } \mathsf{w}_b \wedge Gx \text{ in } \mathsf{w}_c)$

There appear to be a number of problems getting from our regimentation of (1) to our proposed truth-conditions. One is that, where (R1) contains two belief operators, (T1) contains only one quantifier over doxastic alternatives. Another is that the syntactically free occurrence of x in the second conjunct of (R1) appears to correspond to a bound variable in (T1).

The fact that a syntactically free pronoun in (R1) corresponds to a bound variable in (T1) is familiar from other cases of 'cross-clausal anaphora':

(12) A man is walking in the park, and he is whistling a cantata.

(13) If a farmer owns a donkey, he feeds it.

To see the difficulty posed by sentence (12), for example, imagine that we regiment it as:

(12R) $\exists x Mx \wedge Wx$

Since (12) seems to say that there is x such that x is a man walking in the park and x is whistling a cantata, it would be natural to think of its truth-conditions as instead corresponding to something like:

(12T) $\exists x(Mx \wedge Wx)$

Thus, the question arises as to how we provide a semantics for (12R) that makes it equivalent to (12T).

There are, of course, various solutions to this last problem. Dynamic semantics, for example, is one well-known family of approaches to issues of cross-clausal anaphora, and the theory developed below is a variant of the dynamic

[12] We write w_a for the result of applying function w to individual a.

[13] Here, W^A is the set of all indexing functions, i.e. the set of all functions from A to W.

theory known as Dynamic Predicate Logic (DPL) (Groenendijk and Stokhof, 1991). Our variant of DPL allows the indefinite *a witch* in the first conjunct of (1) to control the interpretation of the pronoun *her* in the second conjunct. This is achieved by allowing a clause of the form *a believes that something is F* to update a certain type of relation that we call an *accessibility relation.* While the term *accessibility relation* is typically used in modal logic to refer to a binary relation between worlds, we appropriate it here to refer to relations that hold between possible worlds and pairs consisting of an indexing function and a variable assignment.

Before we present the semantics, I should mention that I suspect that the basic insight guiding our proposal can be implemented in frameworks other than DPL, for example, in alternative versions of dynamic semantics or in a suitably sophisticated static semantic framework.[14] I have chosen to give a concrete implementation of the proposal in DPL because that theory is reasonably well-known and comparatively simple. But my main aim in what follows is to show that *there is some way* of compositionally implementing our relationist proposal, rather than to argue for one particular way of doing so.

Our approach is to translate sentences of English into a formal language, and then state our semantics for the formal language. Given a non-empty set of agents A, we assume a language $\mathcal{L}_A$. (In the intended application, A would be the set of all actual and possible agents.) The vocabulary of this language consists of n-ary relation symbols, variables, $\neg$, $\wedge$, $\exists x$, and, for each $a \in A$, a belief operator $\mathcal{B}_a$ (the other logical symbols may defined in the usual way, e.g. $(\phi \vee \psi) =_{df} \neg(\neg\phi \wedge \neg\psi)$, $\forall x\phi =_{df} \neg\exists x\neg\phi$). The definition of the formulas of $\mathcal{L}_A$ can be gleaned from the recursive semantics below. As before, we translate (1) into this language as $\mathcal{B}_b\exists xFx \wedge \mathcal{B}_c Gx$.

Definition 1. A *model* for $\mathcal{L}_A$ is a tuple $M = (W, D, B, I)$ where:

(1) W is a non-empty set, whose elements we call *worlds*,

(2) D is a non-empty set whose elements we call *individuals*,

(3) B is a relation between worlds $w \in W$ and elements $\mathsf{w} \in W^A$ (i.e. $B \subseteq W \times W^A$), and

(4) I is a function which maps an n-ary relation symbol and a world to a subset of D^n.

Regarding relation B in clause (3): Suppose $w \in W$ and $\mathsf{w} \in W^A$. Then the idea is that, in an intended model, $wB\mathsf{w}$ iff the indexing function w is compatible with what the agents in A believe in w.

The definitions that follow are given relative to a fixed model $M = (W, D, B, I)$ for a fixed language $\mathcal{L}_A$.

A variable assignment is a function from the variables of $\mathcal{L}_A$ into D. We use $\mathcal{G}$ to denote the set of variable assignments. If g and h are variable assignments

[14] For other dynamic approaches to cross-clausal anaphora, see Heim (1982) and Kamp (1981). For static alternatives, see Rothschild (2017) and Mandelkern (2022).

and x a variable, we say that h is an *x-variant* of g, $h[x]g$, iff for all variables y other than x, $h(y) = g(y)$.

Definition 2. A binary relation $R \subseteq W \times (W^A \times \mathcal{G})$ is an *accessibility relation* iff: (i) $wR(\mathsf{w}, g')$ only if $wB\mathsf{w}$, and (ii) if $wB\mathsf{w}$, then $wR(\mathsf{w}, g)$, for some g. We let $R(w) = \{(\mathsf{w}, g) : wR(\mathsf{w}, g)\}$

Definition 3. Given an accessibility relation R, a variable assignment h, and a variable x, let R^h_x be the variable assignment defined as follows: for any w and any (w, g), $wR^h_x(\mathsf{w}, g)$ iff there is a k such that:

(1) $wR(\mathsf{w}, k)$, and

(2) $g[x]k$ and $g(x) = h(x)$.

While this last definition plays no role in our discussion of (1), it is used below in the clause for the existential quantifier in order to secure satisfactory results for *de re* attitude ascriptions. We leave it to the reader to verify that R^h_x, so defined, is an accessibility relation, given that R is.

Definition 4. We now define the semantic value of a formula ϕ relative to: an indexing function $\mathsf{w} \in W^A$, an agent $a \in A$, a pair of variable assignments $g, g' \in \mathcal{G}$, and a pair of accessibility relations R, R'.

(1) $[\![Fx_1, ..., x_n]\!]^{\mathsf{w},a,g,g',R,R'} = 1$ iff $g = g'$, $R = R'$, and $(g(x_1), ..., g(x_n)) \in I(F, \mathsf{w}_a)$

(2) $[\![\neg\phi]\!]^{\mathsf{w},a,g,g',R,R'} = 1$ iff $g = g'$, $R = R'$, and there is no h and Q such that $[\![\phi]\!]^{\mathsf{w},a,g,h,R,Q} = 1$

(3) $[\![\phi \wedge \psi]\!]^{\mathsf{w},a,g,g',R,R'} = 1$ iff there is an h and a Q such that $[\![\phi]\!]^{\mathsf{w},a,g,h,R,Q} = 1$ and $[\![\phi]\!]^{\mathsf{w},a,h,g',Q,R'} = 1$

(4) $[\![\exists x\phi]\!]^{\mathsf{w},a,g,g',R,R'} = 1$ iff there is an h such that $h[x]g$ and $[\![\phi]\!]^{\mathsf{w},a,h,g',R^h_x,R'} = 1$

(5) $[\![\mathcal{B}_b\phi]\!]^{\mathsf{w},a,g,g',R,R'} = 1$ iff $g = g'$ and:

  (a) for all $(\mathsf{w}', h) \in R(\mathsf{w}_a)$, there is an h' and a Q such that $[\![\phi]\!]^{\mathsf{w}',b,h,h',R,Q} = 1$, and

  (b) for all $(\mathsf{w}', h) \in R'(\mathsf{w}_a)$, there is an h' such that $(\mathsf{w}', h') \in R(\mathsf{w}_a)$ and such that, for some Q, $[\![\phi]\!]^{\mathsf{w}',b,h',h,R,Q} = 1$

The first four clauses here are essentially the standard DPL clauses lifted into our system. The one exception is the clause for the existential quantifier, but even here the only difference is the fact that the 'input' relation for evaluating the embedded clause ϕ is updated to R^h_x. The final clause is new, since DPL is defined over a language that lacks belief operators. It is in this clause that our principal innovation arises; this innovation can be seen at work in the proof of **Proposition 1** below.

Despite the fact that we are working in a dynamic system, our real interest is assigning truth-conditions to sentences like (1). So we need to add to the foregoing a definition of *truth at a world*. We may do this as follows. Given a world w, let w^w be the element of W^A such that $\mathsf{w}^w_a = w$, for all $a \in A$ (so w^w is a constant function mapping all elements of A to w). Then we can define *truth at a world* as follows:

Definition 5. A sentence ϕ is *true at a world* w iff for any g, any $a \in A$, and any R, there is a g' and an R' such that $[\![\phi]\!]^{\mathsf{w}^w,a,g,g',R,R'} = 1$.

In order to show that the foregoing theory assigns to (1) our proposed truth-conditions, it will help to first establish the following lemma:

Lemma 1. $[\![\exists xFx]\!]^{\mathsf{w},a,g,g',R,R'} = 1$ iff $g'[x]g$, $R' = R^g_x$, and $g'(x) \in I(F, \mathsf{w}_a)$

Proof. By the clauses for the existential quantifier and for atomic formulas, we have that:

$[\![\exists xFx]\!]^{\mathsf{w},a,g,g',R,R'} = 1$ iff

there is an h such that $h[x]g$ and $[\![Fx]\!]^{\mathsf{w},a,h,g',R^h_x,R'} = 1$ iff

there is an h such that $h[x]g$ and $R^h_x = R'$ and $h = g'$ and $g'(x) \in I(F, \mathsf{w}_a)$ iff

$g'[x]g$ and $R^{g'}_x = R'$ and $g'(x) \in I(F, \mathsf{w}_a)$.

□

Finally, our principal claim:

Proposition 1. $(\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx)$ is true at a world w iff for all indexing functions $\mathsf{w} \in W^A$, if $wB\mathsf{w}$, then there is an $o \in D$ such that $o \in I(F, \mathsf{w}_b)$ and $o \in I(G, \mathsf{w}_c)$.

Proof. Left-to-Right: Suppose $(\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx)$ is true at w. Then for any g, a, and R, there is a g' and R' such that: $[\![\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx]\!]^{\mathsf{w}^w,a,g,g',R,R'} = 1$. So let g, a, and R be arbitrary, and let g' and R' be such that $[\![\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx]\!]^{\mathsf{w}^w,a,g,g',R,R'} = 1$. It follows from this and the clause for conjunction that there is a k and a Q such that:

(i) $[\![\mathcal{B}_b\exists xFx]\!]^{\mathsf{w}^w,a,g,k,R,Q} = 1$ and

(ii) $[\![\mathcal{B}_cGx]\!]^{\mathsf{w}^w,a,k,g',Q,R'} = 1$.

We first establish that claim (i), together with the clause for the belief operator and **Lemma 1**, implies the following:

(A) for all $(\mathsf{w}, h) \in Q(\mathsf{w}^w_a)$, $h(x) \in I(F, \mathsf{w}_b)$.

To see this, first note that, given claim (i) and clause (5b), we have:

for all $(\mathsf{w}, h) \in Q(\mathsf{w}^w_a)$, there are h' and Q' such that $(\mathsf{w}, h') \in R(\mathsf{w}^w_a)$ and $[\![\exists xFx]\!]^{\mathsf{w},b,h',h,R,Q'} = 1$.

Given **Lemma 1**, this implies:

> for all $(\mathsf{w}, h) \in Q(\mathsf{w}_a^w)$, there are h' and Q' such that $(\mathsf{w}, h') \in R(\mathsf{w}_a^w)$ and $h[x]h'$ and $R_x^h = Q'$ and $h(x) \in I(F, \mathsf{w}_b)$.

And this, in turn, implies:

> for all $(\mathsf{w}, h) \in Q(\mathsf{w}_a^w)$, $h(x) \in I(F, \mathsf{w}_b)$

which is just claim (A).

We now show that claim (ii), together with the clauses for the belief operator and atomic formulas, implies the following:

(B) for all $(\mathsf{w}, h) \in Q(\mathsf{w}_a^w)$, $h(x) \in I(G, \mathsf{w}_c)$.

To see this, note that given clause (5a), claim (ii) holds only if:

> for all $(\mathsf{w}, h) \in Q(\mathsf{w}_a^w)$ there exists h' and Q' such that $[\![Gx]\!]^{\mathsf{w},c,h,h',Q,Q'} = 1$.

Given the clause for atomics, this holds iff:

> for all $(\mathsf{w}, h) \in Q(\mathsf{w}_a^w)$ there exists h' and Q' such that $h = h', Q = Q'$, and $h'(x) \in I(G, \mathsf{w}_c)$.

And this implies:

> for all $(\mathsf{w}, h) \in Q(\mathsf{w}_a^w)$, $h(x) \in I(G, \mathsf{w}_c)$

which is just claim (B).

Claims (A) and (B) yield our '*Left-to-Right*' result as follows. Suppose $wB\mathsf{w}$. Note that $w = \mathsf{w}_a^w$, so we are supposing that $\mathsf{w}_a^w B\mathsf{w}$. Since Q is an accessibility relation, there is an h such that $(\mathsf{w}, h) \in Q(\mathsf{w}_a^w)$. So from (A) we have that $h(x) \in I(F, \mathsf{w}_b)$; and from (B) we have that $h(x) \in I(G, \mathsf{w}_c)$. So there is an $o \in D$ such that $o \in I(F, \mathsf{w}_b)$ and $o \in I(G, \mathsf{w}_c)$, for $h(x)$ is such an o.

Right-to-Left: Suppose that for all indexing functions $\mathsf{w} \in W^A$, if $wB\mathsf{w}$, then there is an $o \in D$ such that $o \in I(F, \mathsf{w}_b)$ and $o \in I(G, \mathsf{w}_c)$. We show that $(\mathcal{B}_b \exists x Fx \wedge \mathcal{B}_c Gx)$ is true at w. To show this, we need to show that for any g, a, and R, there is a g' and an R' such that $[\![\mathcal{B}_b \exists x Fx \wedge \mathcal{B}_c Gx]\!]^{\mathsf{w}^w,a,g,g',R,R'} = 1$. So let g, a, and R be arbitrary. We are going to show that g is such a g' and that we have an appropriate R' be defining it as follows: for any (w, h):

(i) $(\mathsf{w}, h) \in R'(\mathsf{w}_a^w)$ iff there is an h' s.t. $(\mathsf{w}, h') \in R(\mathsf{w}_a^w)$ and $h[x]h'$ and $h(x) \in I(F, \mathsf{w}_b)$ and $h(x) \in I(G, \mathsf{w}_c)$.

(ii) for any $w' \neq \mathsf{w}_a^w$, $(\mathsf{w}, h) \in R'(w')$ iff $w'B\mathsf{w}$.

We first need to show that R' so defined is indeed an accessibility relation. For this, we need to establish two things:

(i) For any w' and (w, j), if $(\mathsf{w}, j) \in R'(w')$, then $w'B\mathsf{w}$.

(ii) For any w' and w, if $w'B\mathsf{w}$, then for some h, $(\mathsf{w}, h) \in R'(w')$.

For (i), suppose $(\mathsf{w}, j) \in R'(w')$. There are two cases: either (a) $w' = \mathsf{w}^w_a$ or (b) $w' \neq \mathsf{w}^w_a$.

For (a): Suppose $w' = \mathsf{w}^w_a$. So $(\mathsf{w}, j) \in R'(\mathsf{w}^w_a)$. Then by the definition of R', there is a j' such that $(\mathsf{w}, j') \in R(\mathsf{w}^w_a, g)$. Since R is an accessibility relation, $\mathsf{w}^w_a B\mathsf{w}$, which means $w'B\mathsf{w}$, since $\mathsf{w}^w_a = w'$.

For (b): Suppose $w' \neq \mathsf{w}^w a$. Then since $(\mathsf{w}, j) \in R'(w')$, it follows by the definition of R' that $w'B\mathsf{w}$.

For (ii), suppose $w'B\mathsf{w}$. We need to show that there is a variable assignment h such that $(\mathsf{w}, h) \in R'(w')$. Again, we have the same two cases: either (a) $w' = \mathsf{w}^w_a$ or (b) $w' \neq \mathsf{w}^w_a$.

For (a): Since $\mathsf{w}^w_a B\mathsf{w}$, it follows from our '*Right-to-Left*' hypothesis that there is an $o \in D$ such that $o \in I(F, \mathsf{w}_b)$ and $o \in I(G, \mathsf{w}_c)$. So let o' be such an o. Note that since R is a variable assignment and $\mathsf{w}^w_a B\mathsf{w}$, there is a j such that $(\mathsf{w}, j) \in R(\mathsf{w}^w_a)$. Let h be the x-variant of j such that $h(x) = o'$. Then $(\mathsf{w}, j) \in R(\mathsf{w}^w_a)$ and $h[x]j$ and $h(x) \in I(F, \mathsf{w}_b)$ and $h(x) \in I(G, \mathsf{w}_c)$. So by the definition of R', $(\mathsf{w}, h) \in R'(\mathsf{w}^w_a)$.

For (b): Suppose $w' \neq \mathsf{w}^w_a$. Note that in this case for any h, $(\mathsf{w}, h') \in R'(w')$ iff $w'B\mathsf{w}$. Thus, since $w'B\mathsf{w}$, any variable assignment will be an h such that $(\mathsf{w}, h) \in R'(w')$.

We now show that $[\![\mathcal{B}_b \exists x Fx \wedge \mathcal{B}_c Gx]\!]^{\mathsf{w}^w, a, g, g, R, R'} = 1$. Given the clause for conjunction, this means showing that there is an h and a Q such that:

(i) $[\![\mathcal{B}_b \exists x Fx]\!]^{\mathsf{w}^w, a, g, h, R, Q} = 1$, and

(ii) $[\![\mathcal{B}_c Gx]\!]^{\mathsf{w}^w, a, h, g, Q, R'} = 1$.

We show that g is such an h and R' such a Q. That is, we show:

(i′) $[\![\mathcal{B}_b \exists x Fx]\!]^{\mathsf{w}^w, a, g, g, R, R'} = 1$, and

(ii′) $[\![\mathcal{B}_c Gx]\!]^{\mathsf{w}^w, a, g, g, R', R'} = 1$.

We start with (i′). Given clause (5) this holds iff $g = g$ and

(a) for all $(\mathsf{w}, h) \in R(\mathsf{w}^w)$, there is an h' and a Q' such that $[\![\exists x Fx]\!]^{\mathsf{w}^w, a, h, h', R, Q'} = 1$, and

(b) for all $(\mathsf{w}, h) \in R'(\mathsf{w}^w_a)$, there is a h' and a Q' such that $(\mathsf{w}, h') \in R(\mathsf{w}^w_a)$ and $[\![\exists x Fx]\!]^{\mathsf{w}, b, h', h, R, Q'} = 1$.

For (a): Let (w, h) be any element of $R(\mathsf{w}^w_a)$. We need to show that there is an h' and a Q' such that $[\![\exists x Fx]\!]^{\mathsf{w}^w, a, h, h', R, Q'} = 1$. Given **Lemma 1**, this means that we need to show that there is an h' and a Q' such that $h'[x]h$ and

$R_x^{h'} = Q'$ and $h'(x) \in I(F, \mathsf{w}_b)$. Since $R_x^{h'}$ is a Q' such that $R_x^{h'} = Q'$, it will suffice to show that there is an h' such that $h'[x]h$ and $h'(x) \in I(F, \mathsf{w}_b)$.

Since R is an accessibility relation and $(\mathsf{w}, h) \in R(\mathsf{w}_a^w)$, it follows that $wB\mathsf{w}$. It thus follows from our '*Right-to-Left*' hypothesis that there is an $o \in D$ such that $o \in I(F, \mathsf{w}_b)$ and $o \in I(G, \mathsf{w}_c)$. Let o' be such an o, and let h' be the x-variant of h such that $h'(x) = o'$. Then $h'[x]h$ and $h'(x) \in I(F, \mathsf{w}_b)$, since $o' \in I(F, \mathsf{w}_b)$.

For (b): Suppose $(\mathsf{w}, h) \in R'(\mathsf{w}_a^w)$. We need to show that there is an h' and a Q' such that $(\mathsf{w}, h') \in R(\mathsf{w}_a^w)$ and $[\![\exists x F x]\!]^{\mathsf{w}, b, h', h, R, Q'} = 1$.

Since $(\mathsf{w}, h) \in R'(\mathsf{w}_a^w)$, it follows from the definition of R' that there is a j such that $(\mathsf{w}, j) \in R(\mathsf{w}_a^w)$ and $h[x]j$ and $h(x) \in I(F, \mathsf{w}_b)$ and $h(x) \in I(G, \mathsf{w}_c)$. We show that j and R_x^h are the needed h' and Q'. That is, we show that $(\mathsf{w}, j) \in R(\mathsf{w}_a^w)$ and $[\![\exists x F x]\!]^{\mathsf{w}, b, j, h, R, R_x^h} = 1$. Since we have that $(\mathsf{w}, j) \in R(\mathsf{w}^w a)$, it remains to show that $[\![\exists x F x]\!]^{\mathsf{w}, b, j, h, R, R_x^h} = 1$. By **Lemma 1**, this holds iff $R_x^h = R_x^h$ and $h[x]j$ and $h(x) \in I(F, \mathsf{w}_b)$. This holds since $h[x]j$ and $h(x) \in I(F, \mathsf{w}_b)$.

We now show (ii′), $[\![\mathcal{B}_c G x]\!]^{\mathsf{w}^w, a, g, g, R', R'} = 1$. Given the clause for belief, we need to show that $g = g$ and:

(a) for all $(\mathsf{w}, h) \in R'(\mathsf{w}_a^w)$, there is an h' and a Q' such that $[\![Gx]\!]^{\mathsf{w}, c, h, h', R', Q'} = 1$, and

(b) for all $(\mathsf{w}, h) \in R'(\mathsf{w}^w a)$, there is a h' and a Q' such that $(\mathsf{w}, h') \in R'(\mathsf{w}_a^w)$ and $[\![Gx]\!]^{\mathsf{w}, c, h', h, R', Q'} = 1$.

Let $(\mathsf{w}, h) \in R'(\mathsf{w}^w a)$. Then we may establish both (a) and (b) by showing that h and R' are such that $[\![Gx]\!]^{\mathsf{w}, c, h, h, R', R'} = 1$. Given the clause for atomics, this will hold iff $h = h$, $R' = R'$ and $h(x) \in I(G, \mathsf{w}_c)$. Since $(\mathsf{w}, h) \in R'(\mathsf{w}_a^w)$, it follows from the definition of R' that $h(x) \in I(G, \mathsf{w}_c)$.

□

4 Conclusion

I have argued that our relationist account of intentional identity avoids many of the problems facing non-relationist accounts (§2, §3.2). I have also argued that there is direct motivation for relationism, arising out of pairs of cases like the Connected and Unconnected Cases (§3.1). Furthermore, once we adopt the relationist perspective, a number of things seem to fall neatly into place. We obtain a natural account of the truth-conditions of sentences like (1) (§3.2), and get a cleaner separation between the semantics and metasemantics of intentional identity (§3.3). Finally, I have just now been arguing that we can provide a compositional semantic theory that assigns to sentences like (1) our proposed truth-conditions. Our theory extends a familiar theory of cross-clausal anaphora, and thus brings the phenomenon of intentional identity into closer dialogue with contemporary formal semantics.

Where to go from here? Intentional identity is but one species of a broader genus and it natural to wonder whether the approach taken here might provide the basis of a more general theory of intensional anaphora. Consider, for example, the phenomenon of modal subordination (Roberts, 1989), which may be illustrated by the following example:

(14) A wolf might come in. It would eat you first.

As with intentional identity, the underlined pronoun here is somehow anaphoric on the underlined indefinite description, despite the fact that pronoun and description lie in the scope of different intensional operators. A similar phenomenon arises again in connection with counterfactual attitude reports:

(15) Sue believes that a senator stole the funds, and she wishes that he had never stolen anything.[15]

Given the apparent similarity between these three cases, it would be natural to seek a general account that covered all of them. My hope is that the approach taken above may form the basis of such an account, but that is something I must leave as a matter for future inquiry.

References

Blumberg, Kyle. 2018. "Counterfactual Attitudes and the Relational Analysis." *Mind* 127:521–546.

Braun, David. 2012. "Hob, Nob, and Mythical Witches." In Joseph Keim Campbell, Michael O'Rourke, and Harry S. Silverstein (eds.), *Reference and Referring*, 149–87. Cambridge, MA: MIT Press.

Burge, Tyler. 1979. "Sinning Against Frege." *Philosophical Review* 88:398 – 432.

Cooper, Robin. 1979. "The Interpretation of Pronouns." In F. Heny and H. Schnelle (eds.), *Syntax and Semantics 10*, 61–92. New York: Academic Press.

Cumming, Samuel. 2014. "Indefinites and Intentional Identity." *Philosophical Studies* 168:371–395.

Edelberg, Walter. 1986. "A New Puzzle about Intentional Identity." *Journal of Philosophical Logic* 15:1–25.

—. 1992. "Intentional Identity and the Attitudes." *Linguistics and Philosophy* 15:561–596.

[15] Counterfactual attitudes like wishing and imagining are discussed in Ninan (2008, 2012, 2013), Yanovich (2011), Maier (2015), Blumberg (2018), and Pearson (2018), though pronominal anaphora is not often discussed in this literature. See also Edelberg (2006).

—. 1995. "A Perspectivalist Semantics for the Attitudes." *Noûs* 29:316–342.

—. 2006. "Intrasubjective Intentional Identity." *Journal of Philosophy* 103:481–502.

Elbourne, Paul. 2005. *Situations and Individuals.* MIT Press.

Evans, Gareth. 1977. "Pronouns, Quantifiers, and Relative Clauses (l)." *Canadian Journal of Philosophyournal of philosophy* 7:467–536.

Geach, Peter T. 1967. "Intentional Identity." *The Journal of Philosophy* 64:627–632.

Glick, Ephraim. 2012. "A Modal Approach to Intentional Identity." *Noûs* 46:386–399.

Groenendijk, Jeroen and Stokhof, Martin. 1991. "Dynamic Predicate Logic." *Linguistics and Philosophy* 14:39–100.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases.* Ph.D. thesis, University of Massachusetts, Amherst.

—. 1990. "E-Type Pronouns and Donkey Anaphora." *Linguistics and Philosophy* 13:137–177.

Hintikka, Jaakko. 1962. *Knowledge and Belief.* Ithaca, N.Y.: Cornell University Press.

Kamp, Hans. 1981. "A Theory of Truth and Semantic Representation." In J.A.G. Groenendijk, T.M.V. Janssen, and M.B.J. Stokhof (eds.), *Formal Methods in the Study of Language*, volume 135, 277–322. Amsterdam: Mathematisch Centrum.

King, Jeffrey C. 1993. "Intentional Identity Generalized." *Journal of Philosophical Logic* 61–93.

Lanier, William. 2013. *Intentional Identity.* Ph.D. thesis, Oxford University.

—. 2014. "Intentional Identity and Descriptions." *Philosophical Studies* 170:289–302.

Maier, Emar. 2015. "Parasitic Attitudes." *Linguistics and Philosophy* 38:205–236.

Mandelkern, Matthew. 2022. "Witnesses." *Linguistics and Philosophy* 45:1091–1117.

Ninan, Dilip. 2008. *Imagination, Content, and the Self.* Ph.D. thesis, MIT.

—. 2012. "Counterfactual Attitudes and Multi-Centered Worlds." *Semantics and Pragmatics* 5:1–57. doi:10.3765/sp.5.5.

—. 2013. "Self-Location and Other-Location." *Philosophy and Phenomenological Research* 87:301–331. doi:http://dx.doi.org/10.1111/phpr.12051.

Pearson, Hazel. 2018. "Counterfactual De Se." *Semantics and Pragmatics* 11.

Putnam, Hilary. 1973. "Meaning and Reference." *Journal of Philosophy* 70:699–711.

Roberts, Craige. 1989. "Modal Subordination and Pronominal Anaphora in Discourse." *Linguistics and Philosophy* 12:683–721.

Rothschild, Daniel. 2017. "A Trivalent Account of Anaphora and Presupposition." In *Proceedings of the 21st Amsterdam Colloquium*, 1–13.

Salmon, Nathan. 1998. "Nonexistence." *Noûs* 32:277–319.

—. 2002. "Mythical Objects." In J. Campbell, M. O'Rourke, and D. Shier (eds.), *Meaning and Truth*. Proceedings of the Eastern Washington University and the University of Idaho Inland Northwest Philosophy Conference on Meaning, Seven Bridges Press.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge, MA: MIT Press.

Yanovich, Igor. 2011. "The Problem of Counterfactual *De Re* Attitudes." *Semantics and Linguistic Theory* 21:56–75.