Dynamic Relationism about Belief*

Dilip Ninan

Tufts University, Medford, MA, USA
dilip.ninan@tufts.edu

Abstract

This essay presents a novel account of intentional identity, the phenomenon illustrated by Geach's 'Hob-Nob' sentence. On the account defended here, such sentences express dyadic belief facts that are not in general reducible to monadic belief facts. We show how these non-reductive truth-conditions are predicted by a version of dynamic semantics which allows sentences to update certain kinds of accessibility relations.

1 Introduction

The problem of intentional identity concerns Geach's infamous 'Hob-Nob' sentence:

(1) [There are no witches, but] Hob believes that a witch blighted Bob's mare, and Nob believes that she killed Cob's sow. [4]

This example raises two questions. First, what are the truth-conditions of (1)? Second, how are those truth-conditions compositionally generated from the meanings of its parts? We begin with a brief discussion of extant answers to the first question, and then suggest taking a rather different approach altogether. On the approach taken here, facts of the form *a believes that* ϕ *and b believes that* ψ are not reducible to facts of the form *c believes that* χ. After briefly explaining and motivating this 'relationist' view, we present a version of dynamic semantics that yields our proposed truth-conditions for (1).

2 Descriptivism

On the intended reading of (1), the pronoun *she* in the second conjunct of (1) is not a deictic pronoun; it is in some sense anaphoric on the indefinite noun phrase *a witch* that occurs in first conjunct. And the intended reading is not one that requires there to be any witches—hence the bracketed conjunct *there are no witches*. But what *does* (1) require for its truth? What are the truth-conditions of this sentence?

It is natural to wonder whether the intended reading of (1) can be paraphrased by replacing the pronoun *she* in the second conjunct with a definite description constructed from the linguistic material found in the first (unbracketed) conjunct. Perhaps (1) is equivalent to (2) or to (3):

(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.

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

*For comments on an earlier version, thanks to an audience at PhLiP 7 (Tarrytown, NY), and to Simon Charlow, Josh Dever, Andy Egan, Matt Mandelkern, and Sarah Moss.

(All of the belief ascriptions here should be read *de dicto*.) While each of these sentences might capture a reading of (1), Geach argued in his original article that there is a reading of (1) on which it is equivalent to neither of these. 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 [2]:

> **Newspaper Case**
>
> A number of animals in Gotham Village have recently died quite unexpectedly. Rumors have begun to circulate that these unfortunate events are due to the machinations of a witch. The local newspaper, the Gotham Star, has picked up on these rumors and reported that the witch has been attacking animals and destroying crops. In reality, there is no such person: 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 article 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.

It is clear from this last detail of the story, that neither (2) nor (3) is true in this scenario. Since Nob has no beliefs about either Hob or Bob, neither (2) nor (3) can be true. But it is widely thought that (1) has a reading on which it is true in this scenario.[1] If that is correct, then neither (2) nor (3) captures the intended reading of (1).

Examples like the Newspaper Case have been taken to show that, on the intended reading, (1) requires that Hob's belief and Nob's belief have a *common causal source* [5, 8].[2] 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 this approach, (1) is 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. [8, 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.

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*

[1] Though see [7] and [1] for some doubts about this.

[2] 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 someone killed Cob's sow.

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, this is not obligatory. 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.[3] 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.

3 Dynamic relationism

Within the framework of possible worlds semantics, a sentence of the form *a believes that* ϕ is said to be true iff in every world w compatible with what a believes, ϕ. Framed in this way, our problem 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? The various answers to this we've considered (e.g. *y is the witch who blighted Bob's mare in w'*) have all proved unsatisfactory.

While there might be other answers to this question worth considering,[4] I suggest that we change course and ask a different question altogether. Suppose we instead ask: what is it for a *pair* of worlds (w, w') to be compatible with what the pair (Hob, Nob) believe? 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'. 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 this condition.

This view will be developed below, but let me pause here to explain what I intend by calling this a *relationist* theory. 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'.[5] 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 *reductionism about dyadic belief* is the view that 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'. *Relationism about dyadic belief* is the negation of reductionism.[6]

It might seem as if reductionism is inescapable, but the approach to (1) taken above suggests how relationism could be true. To see what I mean, let a and b be fixed but arbitrary agents. Let $Dox^w_{a,b}$ be the set of pairs of worlds compatible with what (a, b) believe in w, let $Dox^w_a :=$

[3] We often forget how we formed certain beliefs. I believe that a moose once wandered into a school in Saskatoon, but I forget how I formed this belief. (Did I read it in the paper? Did a friend tell me?)

[4] See, for example, [9], [10], and [5].

[5] We state this definition, and the views to follow, in a species of mathematical English that permits quantification into sentence position.

[6] My use of the term *relationism* is inspired by [3], though my version of relationism is rather different from Fine's.

$\{v : (v, v') \in Dox^w_{a,b}$, for some $v'\}$, and let $Dox^w_b := \{v' : (v, v') \in Dox_{a,b}$, for some $v\}$. Assume that for any $c \in \{a, b\}$, c has precisely the same monadic beliefs in w as they have in w' iff $Dox^w_c = Dox^{w'}_c$. And 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 certain 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, reductionism 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 blighted Bob's mare in v and x killed Cob's sow in v')$\}$.

$Dox^{w'}_{a,b} = \{(v, v') : \exists x$ (x blighted Bob's mare in v) and $\exists y$ (y 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 $Dox^w_a = Dox^{w'}_a$ and $Dox^w_b = Dox^{w'}_b$, which means that a and b each have precisely the same monadic beliefs in w as they have in w'.[7] Thus, we have a difference in the relevant dyadic belief facts despite no difference in the relevant monadic belief facts.

The Newspaper Case can be used to provide intuitive motivation for relationism. For we might take that case to show 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) Hob's belief and Nob's belief have a common causal source (of the appropriate kind). That might be too simplistic, but if *something* like this is right, then it should not be too difficult to construct a pair of scenarios w and w' such that Hob has the precisely the same monadic beliefs in w as he has in w', Nob has the precisely the same monadic beliefs in w as he has in w', but (1) is true in w but not in w'. For suppose that (i) and (ii) both obtain in w and w', while (iii) obtains in w but not in w'. Then it is plausible to think that (1) will be true in w, false in w'. But all that appears to be compatible with Hob and Nob each having precisely the same monadic beliefs in w as they have in w'. For the key difference between w and w' concerns the *causal origins* of Hob's belief or of Nob's belief. And it's plausible to think the causal origins of one of those beliefs could differ across w and w' even if Hob and Nob each have precisely the same monadic beliefs in w as they have in w'. Whether or not Hob's belief and Nob's belief have a common causal source is a fact *external* to their monadic beliefs; thus, the presence or absence of a common cause may vary even as their monadic beliefs are held fixed.

While there is more to be said about these foundational matters, we turn now to develop the relationist approach in more detail. In order to generalize the approach, first note that there is nothing special about pairs of believers:

(6) 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.

[7]To see that $Dox^w_a \subseteq Dox^{w'}_a$, suppose $v \in Dox^w_a$. 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 $(v, v') \in Dox^{w'}_{a,b}$, which means $v \in Dox^{w'}_a$. To see that $Dox^{w'}_a \subseteq Dox^w_a$, suppose $v \in Dox^{w'}_a$. So there is a v' such that there is an x that blighted Bob's mare in v and 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 $(v, u) \in Dox^w_{a,b}$, which means $v \in Dox^w_a$, as desired. The argument that $Dox^w_b = Dox^{w'}_b$ is similar.

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 we use *functions* from individuals to worlds. Given a non-empty set W of worlds and a non-empty set of agents A, let $\mathsf{f} \in W^A$ be a *doxastic alternative for* A.[8] Then if $\{\text{Hob}, \text{Nob}\} \subseteq A$, I propose that (1) is true at a world w iff for all doxastic alternatives $\mathsf{f} \in W^A$, there is an x such that x is a witch who blighted Bob's mare in $\mathsf{f}(\text{Hob})$ and x killed Cob's sow in $\mathsf{f}(\text{Nob})$.

We turn now to the task of constructing a compositional semantics that generates these truth conditions. Our theory is a variant of Dynamic Predicate Logic [6] that 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 from $\mathcal{B}_b \exists x \phi$ to update a certain type of relation that we call an *accessibility relation*. These accessiblity relations hold between points that include variable assignments.

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 definition of the formulas of $\mathcal{L}_A$ can be gleaned from the recursive semantics below. We translate (1) into this language as $(\mathcal{B}_b \exists x Fx \wedge \mathcal{B}_c Gx)$.

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

(1) W is a non-empty set of worlds,

(2) D is a non-empty set of individuals,

(3) B is a relation between worlds $w \in W$ and elements $\mathsf{f} \in W^A$ (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 (3): the idea is that, in an intended model, $wB\mathsf{f}$ iff f 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 G to denote the set of variable assignments. If g and h are variable assignments 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 $R \subseteq (W \times G) \times (W^A \times G)$ is an *accessibility relation* iff: (i) $(w, g)R(\mathsf{f}, g')$ only if $wB\mathsf{f}$, and (ii) if $wB\mathsf{f}$, then for any g, there is a g' s.t. $(w, g)R(\mathsf{f}, g')$. Let $R(w, g) := \{(\mathsf{f}, g') : (w, g)R(\mathsf{f}, g')\}$

Definition 3. We define the semantic value of a formula ϕ relative to: a doxastic alternative $\mathsf{f} \in W^A$, an agent $a \in A$, a pair of variable assignments $g, g' \in G$, and a pair of accessibility relations R, R'.

(1) $[\![Px_1, ..., x_n]\!]^{\mathsf{f},a,g,g',R,R'} = 1$ iff $g = g'$, $R = R'$, and $(g(x_1), ..., g(x_n)) \in I(P, \mathsf{f}a)$

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

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

[8] Where W^A is the set of all functions $\mathsf{f} : A \to W$.

(4) $[\![\exists x\phi]\!]^{\mathsf{f},a,g,g',R,R'} = 1$ iff $R = R'$ and there is an h such that $h[x]g$ and $[\![\phi]\!]^{\mathsf{f},a,h,g',R,R'} = 1$[9]

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

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

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

Given a world w, let f^w be the element of W^A such that $\mathsf{f}^w a = w$, for all $a \in A$. And let R_{max} be the accessibility relation R such that for any world w and function $\mathsf{f} \in W^A$, if $wB\mathsf{f}$, then $(w, g)R(\mathsf{f}, g')$, for any variable assignments g and g'. Then we can define *truth at a world* as follows:

Definition 4. A sentence ϕ is *true at a world* w iff for any g and any $a \in A$, there is a (g', R') such that $[\![\phi]\!]^{\mathsf{f}^w,a,g,g',R,R'} = 1$, where $R = R_{max}$.

This view predicts that (1) has our proposed truth-conditions:

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

Proof-sketch of the left-to-right direction. Suppose $(\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx)$ is true at a world w. From the definition of truth at a world, it follows that there is a g' and and R' such that:

$$[\![(\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx)]\!]^{\mathsf{f}^w,a,g,g',R,R'} = 1.$$

(Here a and g are arbitrary and $R = R_{max}$.) It follows from this and from the clause for $\wedge$ that there is a k and a Q such that:

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

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

From (i) and the clauses for $\mathcal{B}$ (5b), $\exists x$, and atomics, it follows that $g = k$ and that:

(A) for all $(\mathsf{f}, h) \in Q(\mathsf{f}^w a, g)$: $h(x) \in I(F, \mathsf{f}b)$.

Since $g = k$, (ii) is equivalent to $[\![\mathcal{B}_cGx]\!]^{\mathsf{f}^w,a,g,g',Q,R'} = 1$. And given the clauses for $\mathcal{B}$ (5a) and atomics, this implies:

(B) for all $(\mathsf{f}, h) \in Q(\mathsf{f}^w a, g) : h(x) \in I(G, \mathsf{f}c)$.

Claims (A) and (B) together imply:

(C) for all $(\mathsf{f}, h) \in Q(\mathsf{f}^w a, g)$: $h(x) \in I(F, \mathsf{f}b)$ and $h(x) \in I(G, \mathsf{f}c)$.

Now suppose $wB\mathsf{f}$. Note that $w = \mathsf{f}^w a$, so we are supposing that $\mathsf{f}^w aB\mathsf{f}$. Since Q is an accessibility relation, there is an h such that $(\mathsf{f}, h) \in Q(\mathsf{f}^w a, g)$. So let (f, h) be any element of $Q(\mathsf{f}^w a, g)$. So from (C) we have that $h(x) \in I(F, \mathsf{f}b)$ and $h(x) \in I(G, \mathsf{f}c)$. So there is an $o \in D$ such that $o \in I(F, \mathsf{f}b)$ and $o \in I(G, \mathsf{f}c)$, for $h(x)$ is such an o. □

[9]A slightly more complex clause is needed here to handle *de re* belief attributions, but I set this issue aside for simplicity.

References

[1] David Braun. Hob, Nob, and mythical witches. In Joseph Keim Campbell, Michael O'Rourke, and Harry S. Silverstein, editors, *Reference and Referring*, pages 149–87. MIT Press, Cambridge, MA, 2012.

[2] Walter Edelberg. A new puzzle about intentional identity. *Journal of Philosophical Logic*, 15:1–25, 1986.

[3] Kit Fine. *Semantic Relationism.* Blackwell Publishing Ltd, Oxford, 2009.

[4] Peter T Geach. Intentional identity. *The Journal of Philosophy*, 64(20):627–632, 1967.

[5] Ephraim Glick. A modal approach to intentional identity. *Noûs*, 46(3):386–399, 2012.

[6] Jeroen Groenendijk and Martin Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100, 1991.

[7] Jeffrey C King. Intentional identity generalized. *Journal of Philosophical Logic*, pages 61–93, 1993.

[8] William Lanier. Intentional identity and descriptions. *Philosophical Studies*, 170(2):289–302, 2014.

[9] Nathan Salmon. Mythical objects. In J. Campbell, M. O'Rourke, and D. Shier, editors, *Meaning and Truth.* Proceedings of the Eastern Washington University and the University of Idaho Inland Northwest Philosophy Conference on Meaning, Seven Bridges Press, 2002.

[10] Robert van Rooij. *Attitudes and Changing Contexts.* Springer, Dordrecht, 2006.