A Relationist Theory of Intentional Identity

Under what conditions is the 'Hob-Nob' sentence true (Geach, 1967)?

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

According to standard possible worlds semantics, the first conjunct is true just in case in all of Hob's 'belief-worlds', there is a witch x who blighted Bob's mare. Intuitively, the second conjunct entails that in all of Nob's belief-worlds, there is a witch y who killed Cob's sow. Framed in that way, the question now becomes: how must the relevant witches in Nob's belief-worlds be related to the relevant witches in Hob's belief-worlds in order for (1) to be true? This question has turned out to be difficult to answer. Extant approaches either send us on a futile search for a definite description that goes proxy for the pronoun in (1) (van Rooij, 2006; Glick, 2012; Lanier, 2014) or tell us that (1) entails the existence of 'mythical witches' (Salmon, 2002).

But perhaps even framing the problem in this way—in terms of Hob's and Nob's *separate* belief-worlds—was a mistake. Suppose we instead begin by asking what it is for a *pair* of worlds (w_h, w_n) to be compatible with what Hob and Nob believe (with w_h 'indexed' to Hob and w_n 'indexed' to Nob). The answer to this question seems comparatively straightforward: if the Hob-Nob sentence is true, then (w_h, w_n) will be compatible with what Hob and Nob believe only if there is an x such that x is a witch in w_h, x blighted Bob's mare in w_h, and x killed Cob's sow in w_n. This is the key idea behind *relationism* about intentional identity. Instead of taking the central doxastic compatibility relation to be the relation 'w is compatible with what agent a believes', the relationist instead proposes to use the relation 'sequence $(w_1, ..., w_n)$ is compatible with what agents $a_1, ..., a_n$ believe'. Or, on my preferred formulation, with the relation 'f is compatible with what the agents in A believe', where f is a function mapping each agent in A to a world in W (the set of all worlds). On this approach, and given a fixed set of agents A that includes Hob and Nob, (1) is true at a world w iff for all $f \in W^A$ compatible with what the agents in A believe in w, there is an x such that x is a witch in $f(\text{Hob})$, x blighted Bob's mare in $f(\text{Hob})$, and x killed Cob's sow in $f(\text{Nob})$. The resulting theory is *relationist* in the sense that 'dyadic facts' of the form *x believes that ϕ and y believes that ψ* will not in general be reducible to the 'monadic facts' about what x believes together with the 'monadic facts' about what y believes (cf. Fine, 2009).

If the above claim about the truth-conditions of (1) is accepted, we may then ask how those truth-conditions are determined compositionally. Suppose we formalize (1) as follows:

(1′) $\mathcal{B}_{hob}(\exists x Fx) \wedge \mathcal{B}_{nob}(Gx)$

Here Fx translates *x is a witch and x blighted Bob's mare* and Gx translates *x killed Cob's sow*. Evidently, we need the existential quantifier $\exists x$ in the first conjunct to bind the seemingly-free occurrence of x in the second conjunct. Problems of this general type are of course familiar, and various solutions to them have been proposed (e.g. Heim 1982; Groenendijk and Stokhof 1991). But those solutions don't straightforwardly generalize to the present case, since in (1′) the existential quantifier and the variable it seeks to bind sit in the scope of different belief operators.[1]

Fortunately, this problem admits of a solution once we adopt the relationist point of view. We demonstrate this by sketching a relationist modification of the dynamic semantics of Groenendijk et al. (1996). Let a *point* (f, g) be a pair of a function $f \in W^A$ and a variable assignment g (a function that maps each variable to an element of a domain $D \supseteq A$), and let a *state* be a set of

[1] Even the literature on modal subordination (Roberts, 1989; Geurts, 1998) is of limited help here, since the two 'modals' in (1′) have different (and possibly incompatible) modal bases.

points. To say that a point 'sees' a point (f,g) implies that f is compatible with what the agents in A believe. Here, very roughly, is the idea behind our treatment of $(1')$. The first conjunct, $\mathcal{B}_{hob}(\exists x Fx)$, takes a state s and updates it to a new state s' whose points can only see a point (f,g) if $g(x)$ is F in $f(\text{Hob})$. Given a dynamic entry for conjunction, the input state for the second conjunct, $\mathcal{B}_{nob}(Gx)$, is s'. $\mathcal{B}_{nob}(Gx)$ then takes s' and updates it to a new state $s'' \subseteq s'$ whose points can only see a point (f,g) if $g(x)$ is G in $f(\text{Nob})$. But since s'' is a subset of s', the result is that points in s'' can only see a point (f,g) if $g(x)$ is F in $f(\text{Hob})$ and $g(x)$ is G in $f(\text{Nob})$. Given a suitable definition of *truth at a world*, we then get the desired result: $(1')$ is true at a world w iff for all $f \in W^A$ compatible with what the agents in A believe in w, there is an individual $o \in D$ such that o is F in $f(\text{Hob})$ and o is G in $f(\text{Nob})$.

That's a rough sketch of how our approach works; the actual workings of the theory are a bit more complex. In particular, in our semantics, a formula updates a *pair* of a state and a relation, and does so *relative to* an agent $a \in A$. The details are as follows. Assume a relation B between worlds $w \in W$ and functions $f \in W^A$, so that wBf iff f is compatible with what the agents in A believe in w. Points and states are as described above. A binary relation R between world-assignment pairs (w,g) and points (f,g') is an *accessibility relation* iff: (i) for any (w,g) and (f,g'), if $(w,g)R(f,g')$, then wBf, and (ii) for any w and f, if wBf, then there are g and g' such that $(w,g)R(f,g')$. Given an accessiblity relation R, let $R(w,g) = \{(f,g') : (w,g)R(f,g')\}$. The semantic value of a formula relative to an agent a is a function from (state, accessibility relation) pairs to (state, accessibility relation) pairs; the recursive definition of this notion is as follows:

1. $(s,R)[Fx]^a = (\{(f,g) \in s : g(x) \in I(F, fa)\}, R)$[2]
2. $(s,R)[\neg\phi]^a = (s - 1((s,R)[\phi]^a), R)$[3]
3. $(s,R)[\phi \wedge \psi]^a = (s,R)[\phi]^a[\psi]^a$
4. $(s,R)[\exists x\phi]^a = (\{(f,g) : \exists o \in D : (f,g) \in 1((s[x/o],R)[\phi]^a)\}, R)$[4]
5. $(s,R)[B_b\phi]^a = (s', R')$, where:
   (a) $s' = \{(f,g) \in s : \{f' : \exists g' : (f',g') \in 1((R(fa,g),R)[\phi]^b\} = \{f' : \exists g' : (f',g') \in R(fa,g)\}\}$, and
   (b) for all (w,g) and (f',g'), $(w,g)R'(f',g')$ iff:
      - if there is an f s.t. $(f,g) \in s'$ and $f(a) = w$, then $(f',g') \in 1((R(fa,g),R)[\phi]^b)$, and
      - if there is no f s.t. $(f,g) \in s'$ and $f(a) = w$, then $(w,g)R(f',g')$.

Given a world w, let $f^w \in W^A$ be the constant function that maps each $a \in A$ to w. Let R_{max} be the accessibility relation such that for any (w,g) and (f,g'), $(w,g)R_{max}(f,g')$ iff wBf. We say that a formula ϕ is *true at world* w iff for some variable assignment g, $(f^w,g) \in 1(S, R_{max}[\phi]^a)$, where S is the set of all states and a is any agent in A. Given this account of truth-at-a-world, we obtain our desired result: $(1')$ is true at a world w iff for all $f \in W^A$ compatible with what the agents in A believe in w, there is an object $o \in D$ such that $o \in I(F, f(\text{Hob}))$ and $o \in I(G, f(\text{Nob}))$.

[2]We write fa for $f(a)$. $I(F, fa)$ is the extension of monadic predicate F at world fa.

[3]Given a pair (s,R), $1(s,R) = s$.

[4]Given an object $o \in D$, $s[x/o] = \{(f,g) : \exists g' : (f,g') \in s \text{ and } g = g'[x/o]\}$. A more complex clause for $\exists x$ is needed to handle *de re* belief ascriptions, but we set that issue aside here for the sake of simplicity.

References

Fine, Kit. 2009. *Semantic Relationism.* Oxford: Blackwell Publishing Ltd.

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

Geurts, Bart. 1998. "Presuppositions and Anaphors in Attitude Contexts." *Linguistics and Philosophy* 21:545–601.

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.

Groenendijk, Jeroen, Stokhof, Martin, and Veltman, Frank. 1996. "Coreference and Modality." In Shalom Lappin (ed.), *Handbook of Contemporary Semantic Theory*, 179–216. Oxford: Blackwell.

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

Lanier, William. 2014. "Intentional Identity and Descriptions." *Philosophical Studies* 170:289–302.

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

Salmon, Nathan. 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.

van Rooij, Robert. 2006. *Attitudes and Changing Contexts.* Dordrecht: Springer.