Roles, Rigidity, and Quantification in Epistemic Logic*

Wesley H. Holliday and John Perry

Abstract Epistemic modal predicate logic raises conceptual problems not faced in the case of alethic modal predicate logic: Frege's "Hesperus-Phosphorus" problem—how to make sense of ascribing to agents ignorance of necessarily true identity statements—and the related "Hintikka-Kripke" problem—how to set up a logical system combining epistemic and alethic modalities, as well as others problems, such as Quine's "Double Vision" problem and problems of self-knowledge. In this paper, we lay out a philosophical approach to epistemic predicate logic, implemented formally in Melvin Fitting's First-Order Intensional Logic, that we argue solves these and other conceptual problems. Topics covered include: Quine on the "collapse" of modal distinctions; the rigidity of names; belief reports and unarticulated constituents; epistemic roles; counterfactual attitudes; representational vs. interpretational semantics; ignorance of co-reference vs. ignorance of identity; two-dimensional epistemic models; quantification into epistemic contexts; and an approach to multi-agent epistemic logic based on centered worlds and hybrid logic.

1 Introduction

In *Modal Logic for Open Minds*, Johan van Benthem (2010b) remarks on the step from modal propositional logic to modal predicate logic: "it is important to perform this extension also from a practical point of view. Knowing objects like persons,

* Version of May 1, 2013. Forthcoming in *Trends in Logic, Outstanding Contributions: Johan F. A. K. van Benthem on Logical and Informational Dynamics*, eds. A. Baltag and S. Smets (Springer).

Wesley H. Holliday
Department of Philosophy, 314 Moses Hall #2390, University of California, Berkeley, CA 94720, e-mail: wesholliday@berkeley.edu

John Perry
Department of Philosophy, HMNSS Building, Room 1604, 900 University Avenue, University of California, Riverside, CA 92521 e-mail: john.perry@ucr.edu

telephone numbers, or even rules and methods is crucial to natural language and human agency" (124). Indeed, talking about such knowledge, as well beliefs and other cognitive attitudes, is also crucial. In ordinary discourse, we freely combine talk about attitudes with predication, quantification, modals, tense, and other constructions. Here are some examples, not all of which are exactly pieces of ordinary discourse, but all of which seem readily intelligible, together with straightforward formalizations—formalizations that we argue below won't quite do:

(1) Elwood believes that JvB wrote *Modal Logic for Open Minds*.

(1_f) $B_e W(j,m)$

(2) There is someone who Elwood believes wrote *MLOM*.

(2_f) $\exists x B_e W(x,m)$

(3) There is a book that Elwood believes JvB wrote.

(3_f) $\exists y(Bk(y) \wedge B_e W(j,y))$

(4) Elwood believes that JvB is J.F.A.K. van Benthem.

(4_f) $B_e j = j'$

(5) Elwood believes that JvB couldn't have been a computer, but could have been a computer scientist.

(5_f) $B_e(\neg\Diamond C(j) \wedge \Diamond CS(j))$

(6) There is someone who Elwood believes couldn't have been a computer, but could have been a computer scientist.

(6_f) $\exists x B_e(\neg\Diamond C(x) \wedge \Diamond CS(x))$

Syntactically, these formalizations combine predicates, variables, names, identity, quantifiers, alethic modal operators, and epistemic modal operators. The progress of logic has involved seeing how to build on the semantics for earlier items on the list to treat items later on the list. In this paper, we will consider to what extent the addition of *epistemic* operators requires departing from the semantics that works well for the previous items on the list—or whether we can get away with the following.

> **Conservative Approach**: add epistemic modal operators with minimal departures from a base semantics for alethic modal predicate logic, e.g., without changing the semantics of *singular terms* or the nature of *possible worlds*.[2]

There is a difficulty with the Conservative Approach, however, which we will call *the problem of the cognitive fix*. This difficulty has led a number of writers to abandon the Conservative Approach. Some have concluded that the "possible worlds"

[2] We do not mean to suggest that there is a consensus on the proper semantics for alethic modal predicate logic. What we have in mind here is the standard development of Kripke-style semantics for modal predicate logic (see, e.g., Braüner and Ghilardi 2006). To the extent that we support the Conservative Approach so understood, one might expect that epistemic operators could be smoothly introduced into alethic modal predicate logic developed in other ways as well. As another point of qualification, we are not arguing for conservativeness with respect to the question of relational vs. neighborhood semantics for epistemic logic (see Arló-Costa and Pacuit 2006).

needed for epistemic logic differ from those needed for alethic modal logic. Some have concluded that names cannot be treated as "rigid designators" in epistemic logic, as they are in alethic modal logic. Some have supposed that the individuals we talk about in epistemic logic are not quite the same as the ones we talk about in alethic modal logic. In this essay, however, we defend the Conservative Approach.

In §2 we discuss the problem of the cognitive fix, its history, and how we propose to handle it. After introducing the formal framework in §3, in §4 we show how our approach resolves a related problem, the so-called Hintikka-Kripke Problem (Lehmann, 1978; Linsky, 1979; Barnes, 1976) for combined alethic-epistemic modal logic. In §5 we show how we deal with quantification into epistemic contexts, and in §6 we extend the framework to multi-agent epistemic logic. Finally, in §7 we conclude with a speculation about how our semantics for static epistemic predicate logic may lead to new directions in dynamic epistemic predicate logic.

In recent years, there has been a wealth of applications of epistemic logic, mostly using propositional languages. While many applications do not need the bells and whistles of variables and quantifiers, others do. This is the point in the quote above from van Benthem, who in addition to doing pioneering work in modal propositional logic has made notable contributions to modal predicate logic (van Benthem, 1985, 1993, 2010a). As he explains at the end of the chapter on Epistemic Logic in *Modal Logic for Open Minds*, "extending our whole framework to the predicate-logical case is a task that mostly still needs to be done—and who knows, it may be done by you!" (144). This is the kind of encouragement from Johan that has launched so many research projects. While we don't pretend to have completed the task, we hope that something here will spark ideas in Johan that will lead to further progress.

2 The Problem of the Cognitive Fix

Although the problems to be dealt with in this paper fall under the province of both linguists and philosophers, our project is one of philosophical logic, not natural language semantics or pragmatics. One of the projects of the philosophical logician is to design formal languages that are unambiguous, clear, and explicit, but that can also capture important kinds of claims about the world expressible in natural language—and that do so in philosophically illuminating ways. Our question in this paper is whether the philosophical logician can design such a language to formalize a class of claims of special interest to philosophers and epistemologists: claims about what agents believe and know. One of the chief obstacles to such a formal analysis, the problem of the cognitive fix, dates back to at least Frege and Russell.

2.1 Frege, Russell and the Problem of the Cognitive Fix

In their development of ideas and options for the foundations of first-order logic, Frege and Russell were not much concerned about necessity and possibility, what we will call the *alethic modalities*. But they were motivated by puzzles about the way names, variables, and identity work in the context of discourse about knowledge and belief, what we will call the *epistemic modalities*. How does the "cognitive significance" of '$a = a$' differ from that of '$a = b$'? How can George IV know that the author of *Waverley* wrote *Waverley*, yet be ignorant of the fact that Scott wrote *Waverley*, given that Scott is the author of *Waverley*? It seems that an agent can know or believe something about an individual, thought about in one way, or, as we shall say, relative to one "cognitive fix," while not believing the same thing of the same individual—and perhaps even believing the opposite—relative to another cognitive fix.[3] These differing cognitive fixes seem to enter into the truth-conditions of reports about what people know and believe. For example, consider:

(7) Elwood believes that the author of *De Natura Deorum* was a Roman.
(8) Elwood believes that the author of *De Fato* was a Roman.

It seems that (7) might be true, while (8) is false, in spite of the fact that Cicero authored both of the works mentioned. In "On Sense and Reference," Frege (1892) treated the problem with his theory of indirect reference: in epistemic contexts such as these, the embedded sentence and its parts do not have their customary reference, but rather refer to their customary *sense*. Hence (7) and (8) do not really report a relation between Elwood and Cicero, the person designated by the descriptions.

In "On Denoting" Russell (1905) used his theory of descriptions to reach a similar conclusion. For example, (7) is rendered as follows, using intuitive abbreviations:

(9) $B_e\, \exists x(Au(x,DeN) \wedge \forall y(Au(y,DeN) \to x = y) \wedge Rom(x)).$

With Russell's treatment, as with Frege's, (7) and (8) contain no reference to Cicero when properly understood. Since there is no reference to Cicero, we cannot substitute the apparently co-referring descriptions; for they do not actually co-refer. On Russell's account, the descriptions both *denote* Cicero. However, sameness of denotation does not support substitution of terms *salva veritate* in general, but only in "extensional" contexts, unlike belief contexts. Although their philosophical tools were quite different, it is generally agreed that Frege and Russell were recognizing a real phenomenon. At least on one permissible reading, the *de dicto* reading, statements like (7) and (8) do not report relations between Elwood and Cicero, and substitution of the singular terms that designate Cicero may fail to preserve truth.

Russell, however, also allowed a second reading of these sentences, where the quantifier takes wide scope:

(10) $\exists x(Au(x,DeN) \wedge \forall y(Au(y,DeN) \to x = y) \wedge B_e\, Rom(x)).$

[3] Wettstein (1988) uses 'cognitive fix' in a more demanding sense, requiring not merely a way of thinking about an object, but also accurate beliefs about what distinguishes the object from others.

On this reading, (7) and (8) assert a relation between Elwood and Cicero, and the substitution of the descriptions will preserve truth. The reading Russell is getting at here, the *de re* reading, has also been widely, but not universally, recognized. In a de re belief report, a relation is asserted between the believer and the object about which she has a belief, and substitution of co-designative terms preserves truth.

The de re reading seems natural when names are used rather than descriptions:

(11) Elwood believes that Cicero was a Roman.
(12) Elwood believes that Tully was a Roman.

However, it is not at all obvious that this is correct. Names like ‘Cicero’ and ‘Tully’ give rise to the problem with which Frege begins “On Sense and Reference” (1892). The statement ‘Cicero = Cicero’ seems trivial, while the statement ‘Cicero = Tully’ contains valuable information; they have different cognitive significance. If Elwood is a rather desultory student, who has heard of Cicero in Philosophy class, where he was not identified as a Roman, and has heard of Tully in Classics, where he was clearly identified as a Roman, Elwood might not know that Cicero was Tully. For him, ‘Cicero = Tully’ might contain just the information he will need on his mid-term. Before he gets it, it seems that (11) might be false, while (12) is true.

Frege is usually understood to have treated names as having senses that pick out their referents by conditions they uniquely fulfill, what Carnap (1947) was to call “individual concepts.” So for Frege names behave like descriptions in epistemic contexts, referring to their customary senses. In a perfect language, the sense of a proper name would be established by the rules of the language. But Frege realized that for ordinary proper names this may not be so; the relevant sense may be clear from context, perhaps constructed from the speaker’s beliefs or generally accepted beliefs about the object. It is the sense of the proper name, not the individual picked out, that individuates the proposition referred to by the that-clause in a belief report; thus, Frege does not allow for de re belief ascriptions in any straightforward way.

Russell thought that statements like (11) and (12) *would* report de re beliefs, so long as ‘Cicero’ and ‘Tully’ were taken to be “logically proper names.” As his views developed, however, he came to doubt that any ordinary names were logically proper names. To assign a logically proper name, he thought, we have to be acquainted with the thing named, and he came to think that we are only acquainted with our own sense data, the properties and relations that obtain among our sense data, and perhaps our selves. All our beliefs about ordinary objects are de dicto; such objects are not known by acquaintance but only by description (Russell, 1911). For Russell, there are de re beliefs, but not about tables and chairs and other people.[4]

Thus, neither Frege nor Russell would countenance the formalizations $(1f)$ - $(6f)$ of (1) - (6).

[4] It seems that this leaves Russell without a solution to the problem of the cognitive fix, relative to logically proper names. For a discussion of this issue, see Wishon 2012.

2.2 Quine on the Collapse of Modal Distinctions

Perhaps 1953 was the nadir of modal logic in philosophy. Quine argued in his "Three Grades of Modal Involvement" (1953) that the whole enterprise was based on a mistake. His argument basically transfers the problem of the cognitive fix from the epistemic realm to the alethic, concluding that if we must have alethic modalities, they should be limited to the *de dicto*, without quantification into alethic contexts.

Granting—only for the sake of argument—that analyticity made sense, Quine could understand necessity and possibility as properties of sentences, *being analytic* and *being synthetic*. If we think it is analytic, and hence necessary, that philosophers are thoughtful, and we are thoughtful philosophers, we should express this as

(13) 'Philosophers are thoughtful' is necessary.

That's the "first grade" of modal involvement: modality expressed as predicates of sentences. The second grade of modal involvement, to which less thoughtful philosophers sink, blurs the use-mention distinction; they say things like:

(14) Necessarily, the heaviest philosopher in the room is thoughtful;
(15) Possibly, the heaviest object in the room is thoughtful;

or, in logical symbols:

(16) $\Box$(the heaviest philosopher in the room is thoughtful);
(17) $\Diamond$(the heaviest object in the room is thoughtful).

If we restrict the aspiring modal logician to statements like this, we will have allowed her what we might think of as *de dicto* alethic modality. The modal operator blocks the ordinary interpretation of the material inside, much like quotation would.

Nothing in the theory of necessity as analyticity allows us to make sense of the third grade of modal involvement, that is, *de re* alethic modality, as would be required to make sense of "quantifying in":

(18) $\exists x \Box(x$ is thoughtful$)$.

And in fact, Quine thinks that this makes no sense, at least if we retain an orthodox interpretation of predication, identity, names, variables, and quantifiers.

The basic problem with the move to (18), Quine thought, is that it requires the aspiring modal logician to give unequal treatment to singular and general terms. Surely she must insist on *extensional opacity*: "Russell was a human" can be necessary, while "Russell was a featherless biped" is contingent, even if actually all and only humans are featherless bipeds. How we designate a *class* is crucial for the modal status of a sentence. But to make sense of quantification into modal contexts, the modal logician must maintain that it makes sense for an open sentence like

(19) $\Box(x$ is thoughtful$)$

to be true of an object absolutely, not relative to some description or other. For the alethic modalities at least, she must insist that how we designate *individuals* doesn't matter. She should, then, accept *referential transparency* and *substitutivity*:

> If t and t' are co-designative singular terms, substitution of t for t' preserves truth, even in alethic modal contexts.

But combining extensional opacity and referential transparency won't work, or so Quine maintained. He used a version of what is now called "the slingshot" to argue that if we accept both doctrines, then the modalities "collapse": all sentences with the same truth value have the same modal status—all necessary or all contingent.

Assume that

(A) Van Benthem is a human.
(D) Van Benthem is a logician.

are both true, but (A) is necessary while (D) is contingent. So (A) and (D) have different modal statuses. Then it seems we should grant that (B) has the same modal status as (A), and (C) has the same modal status as (D):

(B) $\{x \mid x = \emptyset \ \&$ van Benthem is a human$\} = \{\emptyset\}$.
(C) $\{x \mid x = \emptyset \ \&$ van Benthem is a logician$\} = \{\emptyset\}$.

The only way (B) could be false is if van Benthem were not a human. Assuming (A) is necessary, that's impossible, so (B) must also be necessary. It's sufficient for (C) to be false that van Benthem not be a logician. Assuming (D) is contingent, that's indeed possible, so (C) must also be contingent. But note that (B) and (C) differ only in that co-referential expressions are substituted one for the other on the left side of the identity sign. So by *substitutivity*, the sentences that result from prefixing 'Necessarily' to (B) and (C) must have the same truth value, i.e., (B) and (C) must have the same modal status, and hence so must (A) and (D). The modalities collapse.

Føllesdal (1961) identified a flaw in Quine's argument.[5] The aspiring modal logician need not claim that *all* co-referential singular terms are intersubstitutable in alethic modal contexts, only that some are; and the step from (B) to (C) will only work if the class abstracts on the left sides of the identity signs are among them.

What would make a class of singular terms such that co-referential terms in the class are always intersubstitutable for each other in alethic modal contexts? It would be so if each term is a *rigid designator* in Kripke's (1980) sense, designating the same object in every possible world (where the object exists). But we can also explain why it would be so in a more general way. Consider the following:

(20) $H(\text{Cicero})$;
(21) $\Box H(\text{Cicero})$.

The standard semantics for (20) tells us that it is true if and only if the open sentence '$H(x)$' is true of *the individual designated by* 'Cicero', whether 'Cicero' is treated as a hidden description or a logically proper name. Let us say that if 'Cicero' is a

---

[5] See Perry 2013 for a fuller account of the slingshot and Føllesdal's treatment of it.

modally loyal term, then the semantics for (21) is such that (21) is true *iff* the open sentence '$\Box H(x)$' is true of that same individual. The same goes for $\Box H(\text{Tully})$, assuming 'Tully' is also a modally loyal term designating the same individual. And an analogous point holds for any other sequence of $\Box$'s and $\Diamond$'s applied to $H(\text{Cicero})$. To have the desired loyalty, the individual supplied for each sentence needs to be the same; and no further conditions that can vary with co-referential terms can be brought into the truth-conditions for sentences that are formed by adding operators.

There are at least two ways terms can be modally loyal. Føllesdal's (1961) "genuine names" are loyal because they are not hidden descriptions or some other complex names built up from general terms with descriptive content. Kripke (1980) argued that ordinary names, like 'Cicero' and 'Tully', are genuine in this sense.

A second source of loyalty, which could be called *sequestering*, can be illustrated by Kaplan's (1989) operator 'Dthat'. 'Dthat' converts descriptions into what Kaplan calls "terms of direct reference." On a standard theory of descriptions, 'the ϕ' *denotes* the unique object, if there is one, that is ϕ. According to Kaplan, the singular term 'Dthat(the ϕ)' *directly refers* to the object denoted by 'the ϕ'. Whatever may be suggested by 'directly', the cash value in alethic modal contexts is this: it is the object denoted by 'the ϕ' that needs to satisfy the open sentence that results from replacing all ocurences of 'Dthat(the ϕ)' in the original sentence by the same variable; and this is *all* that 'Dthat(the ϕ)' makes available; in particular, the relevance of the condition of being a ϕ is exhausted in determining the reference and plays no further role in the semantics of modal sentences in which 'Dthat(the ϕ)' occurs.

Genuine names and sequestered descriptions suggest a solution to Quine's argument. Substitutivity holds for some singular terms, but not all. The class abstracts at the beginning of (B) and (C) are not among the singular terms for which substitutivity holds. The class abstracts are not genuine names or sequestered descriptions, but rather complex names whose semantics is based on that of general terms.[6]

[6] What if we amend (B) and (C) so that they begin with sequestered terms? Extending 'Dthat' to class abstracts, we have:

(A) Van Benthem is a human.
(B′) Dthat($\{x \mid x = \emptyset$ & van Benthem is a human$\}) = \{\emptyset\}$.
(C′) Dthat($\{x \mid x = \emptyset$ & van Benthem is a logician$\}) = \{\emptyset\}$.
(D) Van Benthem is a logician.

Could Quine still argue that (A) and (B′), and (C′) and (D), have the same modal status? (D) is clearly contingent; the multi-talented van Benthem could have been a computer scientist. Is (C′) contingent? The result of applying Dthat to a description or class abstract is supposed to be a rigid designator, or more generally, a modally loyal term in the sense defined above. Thus, evaluating

$$\Box\text{Dthat}(\{x \mid x = \emptyset \ \& \text{ van Benthem is a logician}\}) = \{\emptyset\}$$

amounts to checking for every possible world whether '$y = \{\emptyset\}$' is true there, where y is assigned the object that is designated by 'Dthat($\{x \mid x = \emptyset$ & van Benthem is a logician$\}$)' *in the actual world* (or the world of the context of utterance). Since the object designated by 'Dthat($\{x \mid x = \emptyset$ & van Benthem is a logician$\}$)' in the actual world is $\{\emptyset\}$, the check succeeds, so (C′) is necessary. Hence (C′) and (D) do not have the same modal status; so the modalities do not collapse.

2.3 Names in Epistemic Logic

Although it no doubt remains for alethic modal logic to reach its zenith, Quine's attempts to undermine it are but a dim memory of older philosophers. Modally loyal names are ubiquitous, both syntactically simple genuine names, functioning basically like variables except with reference fixed in the Lexicon,[7] and complex but sequestered singular terms, as in Kaplan's (1989) *Demonstratives*.

It is widely held, however, that modally loyal names, understood as they are in alethic logic, do not work for epistemic logic. Alethic modal logic can perhaps ignore the problem of the cognitive fix. But cognition is the stuff that reports of belief and knowledge seem to be about, and the problem won't go away.

The problem is perhaps clearer in the case of sequestered names than in that of genuine names. The beliefs that might motivate a competent speaker to say or write

(22) Dthat(the author of *Satan in the Suburbs*) resigned from UCLA.

might differ from the beliefs that would motivate him to say or write

(23) Dthat(the author of *Marriage and Morals*) resigned from UCLA.

Someone might be in a position to sincerely and confidently utter the second, without having any idea about the first. The features of sequestering that allow the aspiring alethic modal logician to evade Quine's argument do not seem to automatically help the aspiring epistemic logician with the problem of the cognitive fix.

Even with genuine names, the problem of the cognitive fix doesn't go away. When we are talking about Cicero's modal properties, it doesn't seem to matter whether we call him 'Cicero' or 'Tully'. But as we saw above, it seems to matter a lot when we are talking about what people know and believe.

The apparent mismatch between the alethic and epistemic modalities comes out in the contrast between the following:

(24) If it is necessary that Tully was a Roman, then it is necessary that Cicero was a Roman.

(24_f) $\Box Rom(Tully) \rightarrow \Box Rom(Cicero)$

(25) If Elwood believes that Tully was a Roman, then Elwood believes that Cicero was a Roman.

(25_f) $B_e Rom(Tully) \rightarrow B_e Rom(Cicero)$

Under the standard interpretation of alethic modality, (24)/(24_f) is valid. If Tully was a Roman in every possible world, then so was Cicero. But intuitively, (25) is not valid. Perhaps Tully is a Roman in every world compatible with what Elwood believes, for Elwood realizes that Tully was a Roman. But if he doesn't believe that Tully is Cicero, and doesn't believe that Cicero was a Roman, it seems there

---

[7] In what follows we consider approaches that assimilate proper names in natural language to *constants* in a formal modal language. We do not have room to discuss alternative approaches, e.g., that assimilate names to predicates (Burge, 1973) or variables (Cumming, 2008).

are worlds, compatible with what he believes, in which Cicero isn't a Roman. But Cicero *is Tully*. So how can (25) be false? There are two main schools of thought.

One approach is to deny the disanalogy between (24) and (25) and claim that (25) is also valid. If Elwood believes that Tully was a Roman, then he believes that Cicero was too. Epistemic contexts are transparent for names. According to a view in the philosophy of language that we will call the *the Heroic Pragmatic theory*,[8] if one says, "Elwood believes that Tully was a Roman, but Elwood doesn't believe that Cicero was a Roman," then one says something false. But one may nevertheless convey something true; the choice of 'Tully' for the first clause and 'Cicero' for the second may convey a Gricean implicature that is true, namely that Elwood wouldn't use the name 'Cicero' to express his belief. By denying the disanalogy between (24) and (25), the Heroic Pragmatist supports the Conservative Approach of §1 for extending alethic to epistemic modal logic, but we think at too high an intuitive cost.

The other approach, more common among philosophical logicians, accepts the intuitive disanalogy between (24) and (25), and rejects the Conservative Approach. While the syntax of epistemic modal logic may parallel that of alethic modal logic, as (25_f) parallels (24_f), the semantics cannot. According to the second approach, although 'Cicero = Tully' is true in all possible worlds, in order to invalidate (25_f) we must allow 'Cicero = Tully' to be false in some *doxastic possibility*. We will call this *the Special Semantics theory*, since it claims that we must modify our semantics when we move from alethic to epistemic modal logic. Some explain this by claiming that in epistemic contexts, names are not rigid designators (see §4.1). Others explain it by a difference between doxastic/epistemic possibilities and possible worlds.

Both approaches accept an assumption that we reject about how the philosophical logician should translate natural language reports into the modal language.

> **Complement = Operand Hypothesis**: in epistemic logic, as in alethic modal logic, the sentence to which the modal operator is applied—the *operand*—is the (formalization of the) sentence embedded in the 'that'-clause—the *complement sentence*—in the natural language belief/knowledge ascription.

Consider the relation between (24) and (24_f). To form the antecedent of the latter, we attach the alethic operator '$\Box$' to the (formalization of the) sentence embedded in the that-clause of the former, 'Tully was a Roman': $\Box Rom(Tully)$; similarly with the consequent. This is a natural step, and goes with thinking of '$\Box$' as being basically a translation of the natural language phrase, 'it is necessary that' (or 'necessarily').

It is then natural to follow the same procedure with epistemic operators, as we did in going from (25) to (25_f). But in doing so, we build in an assumption that is not part of the formal framework of modal logic. In doing epistemic modal logic, we want to use the apparatus of modal logic to analyze belief reports. It does not follow that the sentence operated upon, in the analysis, needs to be the same as the sentence embedded in the 'that'-clause in the natural language ascription. This assumption, though natural, is on our view incorrect. This assumption, which one can think of as a form of *syntactic conservatism*, has to be rejected in order to maintain, in a way that takes account of intuitions, our approach of *semantic* conservatism.

[8] See, for example, Barwise and Perry 1983 and Soames 1989.

The reason is that natural language reports of belief and knowledge are not fully explicit. On our view, they are not only about objects, but also about the cognitive fixes via which those objects are cognized. One can represent such cognitive fixes with bits of language (names or definite descriptions), mental particulars (ideas or notions), or entities to which these somehow correspond. Frege's *Sinne* and Carnap's individual concepts are examples of the latter. This is the course we follow. But the entities we choose are what we call *agent-relative roles* (cf. Perry 2009).

We assume only that in every case of cognition about an object, there is some relation that holds between the cognizer and the object—the object plays some role vis-a-vis the cognizer—in virtue of which this is the object cognized. Carnap's individual concepts are supposed to provide the same way of thinking about the object for any agent. But in the general case, this requirement is too strong.

Here is a variation of an example from Quine (1956) that makes this clear.

Suppose Ralph knows that the shortest spy is a spy. It doesn't seem that this would make the FBI interested in him. But there is a stronger condition that would make the FBI interested, which we could express by saying: there is someone Ralph knows to be a spy, or, to make the structure a bit more explicit, there is someone such that Ralph knows that *he* is a spy. But surely we can't say that just because Ralph knows that the shortest spy is a spy, because anyone who knows English knows that.

Suppose that Ralph sees his neighbor Ortcutt on the beach, but doesn't recognize him (so he wouldn't say "that man is Ortcutt"). Ralph sees Ortcutt take papers from a bag marked 'CIA' and hand them to an obvious Bolshevik. Here the relevant agent-relative role is *being the man seen*.[9] Ortcutt plays this role relative to Ralph. It is in virtue of his playing this role that Ralph's thought—the one he might express with "that man is a spy"—is *about* Ortcutt. Roles correspond to cognitive fixes.

The police arrive, and they take Ortcutt into custody, but Ralph still doesn't recognize him. After interviewing Ralph, one policeman reports to the other: "Ralph believes that Ortcutt is a spy." The name 'Ortcutt' is associated by the police with a certain role, *being the person identified on the mug sheets as 'Ortcutt'*. But that role is not the one that the police are claiming to play a role in *Ralph*'s cognitive economy. Clearly, their interest in his testimony is due to the fact that Ortcutt was the man Ralph was *seeing*. As we construe the policeman's remark, he says that Ralph had a belief about a certain man, Ortcutt, via the role, *being the man seen*.

As this suggests, on our view roles are "unarticulated constituents" of the reports. That is, there is no morpheme in the report that refers to them, but they figure in the truth-conditions of what is said. This means, in effect, that to carry out a formalization you need a story (context) and not just a sentence. The classic example motivating the idea of unarticulated constituents is an utterance of "It is raining," made in Palo Alto, by way of calling off a tennis match scheduled there (Perry, 1986). No morpheme refers to Palo Alto, but clearly it is the fact that it is raining in Palo Alto that makes the utterance true. While the classic example has proven

[9] Cf. Lewis (1979, 543) on *watching* as a "relation of acquiantance."

quite controversial, no doubt because of the complexity of weather phenomena of all sorts, the same idea was applied to belief reports in Crimmins and Perry 1989.[10]

Thus, on our view, singular terms employed in complement sentences of belief-reports have two functions. Their semantic function is to identify the objects the belief is about. Typically, they also have the pragmatic function of providing evidence about which cognitive fixes, or roles, are relevant. On this much, we agree with the Heroic Pragmatists. We disagree, however, in the way we think cognitive fixes are relevant. The Heroic Pragmatist sees them as triggers for Gricean implicatures. We see them as semantic parameters of the whole belief report. In our formalization of belief-reports, two kinds of information will be fully explicit: the objects the belief are about, and the cognitive fixes on such objects that are involved in the beliefs.

Thus we see virtue in each of the approaches we don't take. On the one hand, the Heroic Pragmatic approach sees correctly that the semantic job of the names in the complement sentence is to refer to the objects believed about. On the other hand, the Special Semantics approach sees correctly that the terms in doxastic operand sentences are not modally loyal (or rigid) in the way that names are in alethic operand sentences. However, each approach is wrong about something too. The Heroic Pragmatic approach does not give cognitive fixes their rightful semantic place. The Special Semantics approach assumes that non-transparency requires non-rigid *names*, whereas in our formalization below, the names will occur in transparent positions, but not within the scope of a doxastic operator; the opacity will be due to thinking of objects via the roles they play, which can differ from world to world.

What, then, should be said about (25) and (25_f)? Let's first look at the sort of case that leads one to think they are invalid. Remember our desultory student Elwood, who knows from his Classics class that Tully was a Roman, and who has heard of Cicero in his Philosophy class, but hasn't learned that he was a Roman there.

On the picture of proper names we have in mind, when people use names they exploit causal/historical networks that support conventions for using the name to refer to objects—typically objects that are the sources of the networks (Perry, 2012). Let $\mathbf{r}_1$ be the role of *being the person at the source of the 'Tully' network exploited by Elwood*, and let $\mathbf{r}_2$ be the role of *being the person at the source of the 'Cicero' network exploited by Elwood*.[11] Although in the actual world, Cicero (= Tully) plays both of these roles, in other worlds compatible with what Elwood believes, different individuals may play these roles—the networks may have different sources.

Let z_1 be what we call a *role-based* (*object*) *variable*, whose interpretation (relative to a current assignment) in any world w is the object that plays role $\mathbf{r}_1$ in w, and

[10] While our strategy is based on the approach of Crimmins and Perry (1989), those authors took cognitive fixes to be *notions* and took notions to be unarticulated constituents of belief-reports. Subsequently Perry has developed an account that is basically similar, but takes cognitive fixes to be *notion-networks*, basically intersubjective routes through notions. Of course, traditionally cognitive fixes have been taken to be individual concepts. We believe that the concept of a *role* provides a general framework into which all of these candidates can be fit.

[11] Cf. Lewis (1979, 542): "If I have a belief that I might express by saying "Hume was noble," I probably ascribe nobility to Hume under the description "the one I have heard of under the name of 'Hume'." That description is a relation of acquaintance that I bear to Hume. This is the real reason why I believe *de re* of Hume that he was noble."

let z_2 be another role-based variable, whose interpretation in any w is the object that plays $\mathbf{r}_2$ in w. Suppose someone utters "Elwood believes that Tully was a Roman, but Elwood doesn't believe that Cicero was a Roman," where the unarticulated role associated with the use of 'Tully' is $\mathbf{r}_1$ and the unarticulated role associated with the use of 'Cicero' is $\mathbf{r}_2$. Then we claim that the truth of the report imposes the following conditions on the actual world and the space of Elwood's doxastic alternatives (where $\boxed{d}\varphi$ says that φ is true in all of Elwood's doxastic alternatives):

(26) $Tully = z_1 \wedge \boxed{d}Rom(z_1) \wedge Cicero = z_2 \wedge \neg\boxed{d}Rom(z_2)$.

In other words, Tully is the person who plays role $\mathbf{r}_1$ in the actual world, and in all worlds v compatible with what Elwood believes, the person who plays $\mathbf{r}_1$ in v is a Roman; but while Cicero is the person who plays role $\mathbf{r}_2$ in the actual world, there is some world u compatible with what Elwood believes such that the person who plays $\mathbf{r}_2$ in u is not a Roman. Since this can happen, (26) can be true. This does not require any possible world compatible with what Elwood believes in which Cicero (= Tully) is not Tully (= Cicero). It simply requires that there are worlds, compatible with what he believes, in which the object that plays role $\mathbf{r}_1$ does not play role $\mathbf{r}_2$.

The examples above involve two kinds of roles exploited in thinking about things: *perceptual* roles and *name-network* roles. Both kinds of roles afford ways of finding out more about an object; they are what Perry calls *epistemic roles* (Perry, 2012).[12] The idea that an agent's believing of an object (*de re*) that it has a property requires, as with (26), that the object be the unique player of some epistemic role for the agent in the actual world, such that in all worlds compatible with the agent's beliefs, the player of the role in that world has the property in question, is closely related to the accounts of *de re* belief proposed by Kaplan (1968) and Lewis (1979, §8), which have been influential in linguistics (cf. Cresswell and von Stechow 1982). When we add the idea that belief reports involving names have such roles as unarticulated constituents, we can make sense of the difference between (24) and (25) (see §4).

Accounts of *de re* belief in the style of Kaplan and Lewis are not without challenges. Ninan (2012) has argued that such accounts do not generalize to handle other attitudes such as *imagination*; and Yalcin (2012) has argued that it is difficult to make such accounts compatible with semantic *compositionality*. We will briefly discuss the first problem in §3.1. For a critical discussion of compositionality taking into account unarticulated constituents, we refer to Crimmins (1992, Ch. 1).

On our view, belief reports get at important cognitive aspects of agents, and keeping track of them helps us understand agents' behavior. First and foremost, in this paper we are interested in using a formal language to describe these aspects of agents, not in doing the formal semantics of natural language belief reports. This is in line with how epistemic logic is used in theoretic computer science and AI, as a tool to model the information that agents have and how they update it, not the meanings of English sentences containing the words 'knows' and 'believes'.

The same can be said of quantified epistemic logic. Although our examples will be drawn from philosophy rather than computer science or AI, notions of de re

[12] Cf. Lewis (1983, 10f) on "relations of epistemic rapport" or "relations of acquaitance."

belief and "double vision" can be applied to computers as well as humans (cf. Belardinelli and Lomuscio 2009); once again, one may be interested in the phenomena themselves, rather than in English talk about them. With this distinction in mind, even the Heroic Pragmatist about belief talk could agree with much of the formal modeling to follow, if he does not disagree with the basic picture of agents acquiring information about objects via the roles these objects play for the agents. For the purpose of modeling agents themselves, whether we take facts about roles to be involved in the semantics or the pragmatics of belief talk is not crucial.

In the next section we introduce the formal framework with which we will implement these ideas, Fitting's (2004; 2006, §5) First-Order Intensional Logic (FOIL).

3 Formal Framework

Our formal language is a slight variant of that of Fitting's (2006, §5) FOIL.[13]

Definition 1 (Language). Fix a set $\{n_1, n_2, \dots\}$ of constant symbols (or *names*); two disjoint sets $\{x_1, x_2, \dots\}$ and $\{r_1, r_2, \dots\}$ of *object variables* and *role variables*; and for every $m \in \mathbb{N}$, a set $\{Q_1^m, Q_2^m, \dots\}$ of *m-place relation symbols*. The *terms t* and *formulas* φ of our language are generated by the following grammar ($i, k \in \mathbb{N}$):

$$t ::= n_i \mid x_i$$
$$\varphi ::= Q_i^k(t_1, \dots, t_k) \mid t = t' \mid \mathsf{P}(t, r_i) \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \boxed{a}\varphi \mid \boxed{d}\varphi \mid \forall x_i \varphi \mid \forall r_i \varphi.$$

We will often replace the subscripts on variables with suggestive letters or words (or drop them altogether) and write out italicized English names in place of $n_1, n_2, \dots$. We define $\exists y\varphi$ as $\neg\forall\neg y\varphi$ for y an object or role variable, and sometimes we use $\Box$ to represent both $\boxed{a}$ and $\boxed{d}$. For the new symbols, we give the following readings:

$\mathsf{P}(t, r_i)$	"t plays role r_i for the agent";
$\boxed{a}\varphi$	"it is alethically necessary that φ";
$\boxed{d}\varphi$	"it is doxastically necessary for the agent that φ".

In §6, we extend the language to describe the beliefs of multiple agents. Of course, one could also introduce an epistemic necessity operator $\boxed{e}$ for knowledge (see Holliday 2013a,b for discussion of additional issues raised by knowledge).

Our models are a standard kind of modal models with constant domains.[14]

[13] There are a few small differences. First, Fitting allows relation symbols Q (but not $=$) to apply to what we call *role variables*—his *intensional variables*—whereas to simplify the definition of the language, we do not. Second, we include constant symbols in the language, whereas for convenience Fitting does not. Third, we have a bimodal language with $\boxed{a}$ and $\boxed{d}$, whereas Fitting has a monomodal language with $\Box$. Finally, where we write $\mathsf{P}(t, r_i)$, Fitting would write $\mathsf{D}(r_i, t)$.

[14] These models are almost the same as those for "contingent identity systems" in Parks 1974 (cf. Hughes and Cresswell 1996, 333) and Priest 2002, §8, but for a few differences: we follow Fitting in allowing F to contain *partial* functions; Parks does not deal with constants; and while Priest does deal with constants, he treats them as *non-rigid*. The differences between our models and

Definition 2 (Constant Domain Models). A (constant domain) model for the language of Definition 1 is a tuple $\mathcal{M} = \langle W, R_a, R_d, D, F, V \rangle$ such that:

1. W is a non-empty set of *worlds*;
2. R_a is a binary *alethic accessibility* relation on W;
3. R_d is a binary *doxastic accessibility* relation on W;
4. D is a non-empty set of *objects*;
5. F is a non-empty set of *roles*, which are partial functions from W to D;
6. V is a *valuation* function such that for a relation symbol Q_i^k and world w, $V(Q_i^k, w)$ is a set of k-tuples of objects; and for a name n_i, $V(n_i, w)$ is an object, for which we assume the following:
   (a) R_a*-rigidity of names* – for all $w, v \in W$, if wR_av, then $V(n_i, w) = V(n_i, v)$;
   (b) R_d*-rigidity of names* – for all $w, v \in W$, if wR_dv, then $V(n_i, w) = V(n_i, v)$.

For $w \in W$, the pair $\mathcal{M}, w$ is a *pointed model*.

Instead of stating (a) and (b), we could simply require that for all names n_i and worlds $w, v \in W$, $V(n_i, w) = V(n_i, v)$. However, for the purposes of our discussion in §4, it helps to distinguish (a) and (b); for according to a common view, we should assume only the R_a-rigidity of names, not the R_d-rigidity of names.

Functions from W to D are traditionally thought of as *individual concepts*, whereas we think of the partial functions from W to D in F as representing *agent-relative roles*.[15] The distinction is not just terminological. The role view leads us to reject constraints on F that have been proposed assuming the individual concept view (in §5), and it leads to important differences in the multi-agent case (in §6).

It would be natural to assume that R_a is a *reflexive* relation—perhaps even an equivalence relation—while R_d is a *serial* relation. But nothing here will turn on properties of the relations, so we prefer to define the most general model classes.

The next step in introducing the semantics is to define variable assignments.

Definition 3 (Variable Assignment). Given a model $\mathcal{M}$, a *variable assignment* μ maps each object variable x_i to some $\mu(x_i) \in D$ and each role variable r_i to some $\mu(r_i) \in F$. For $d \in D$, let $\mu[x_i/d]$ be the assignment such that $\mu[x_i/d](x_i) = d$ and for all other x_j, $\mu[x_i/d](x_j) = \mu(x_j)$. For $f \in F$, $\mu[r_i/f]$ is defined analogously.

We can now define the interpretations of the two types of terms in our language.

Definition 4 (Interpretation of Terms). The interpretation $[t]_{\mathcal{M},w,\mu}$ of a term t in model $\mathcal{M}$ at world w with respect to assignment μ is an object given by:

$$[n_i]_{\mathcal{M},w,\mu} = V(n_i, w);$$
$$[x_i]_{\mathcal{M},w,\mu} = \mu(x_i).$$

Fitting's (2004; 2006) are that we deal with constants, and Fitting defines V so that predicates can apply not only to elements of D, but also to elements of F (cf. Hughes and Cresswell 1996, 344ff).

[15] We are not suggesting that all there is to a role is a partial frunction from W to D; but such a function captures an important aspect of a role, namely the players of it across worlds.

It follows from Definition 4 and parts (a) and (b) of Definition 2 that

for all n_i and $w, v \in W$, if $wR_a v$, then $[n_i]_{\mathscr{M},w,\mu} = [n_i]_{\mathscr{M},v,\mu}$;
for all n_i and $w, v \in W$, if $wR_d v$, then $[n_i]_{\mathscr{M},w,\mu} = [n_i]_{\mathscr{M},v,\mu}$.

This is the sense in which a name n_i is a *rigid designator*. Given the first point, we can call n_i an "alethically rigid designator," and given the second point, we can call n_i a "doxastically rigid designator." We will return to these points in §4.

Finally, we are ready to state the truth definition for our version of FOIL.

Definition 5 (Truth). Given a pointed model $\mathscr{M}, w$, an assignment μ, and a formula φ, we define $\mathscr{M}, w \vDash_\mu \varphi$ (φ is true in $\mathscr{M}$ at w with respect to μ) as follows:

$$\begin{array}{lll}
\mathscr{M}, w \vDash_\mu Q_i^k(t_1, \dots, t_k) & \text{iff} & \langle [t_1]_{\mathscr{M},w,\mu}, \dots, [t_k]_{\mathscr{M},w,\mu} \rangle \in V(Q_i^k, w); \\
\mathscr{M}, w \vDash_\mu t = t' & \text{iff} & [t]_{\mathscr{M},w,\mu} = [t']_{\mathscr{M},w,\mu}; \\
\mathscr{M}, w \vDash_\mu \mathsf{P}(t, r_i) & \text{iff} & \mu(r_i) \text{ is defined at } w \text{ and } [t]_{\mathscr{M},w,\mu} = \mu(r_i)(w); \\
\mathscr{M}, w \vDash_\mu \neg\varphi & \text{iff} & \mathscr{M}, w \nvDash_\mu \varphi; \\
\mathscr{M}, w \vDash_\mu (\varphi \wedge \psi) & \text{iff} & \mathscr{M}, w \vDash_\mu \varphi \text{ and } \mathscr{M}, w \vDash_\mu \psi; \\
\mathscr{M}, w \vDash_\mu \forall x_i \varphi & \text{iff} & \text{for all } d \in D,\ \mathscr{M}, w \vDash_{\mu[x_i/d]} \varphi; \\
\mathscr{M}, w \vDash_\mu \forall r_i \varphi & \text{iff} & \text{for all } f \in F,\ \mathscr{M}, w \vDash_{\mu[r_i/f]} \varphi; \\
\mathscr{M}, w \vDash_\mu \boxed{a}\varphi & \text{iff} & \text{for all } v \in W, \text{ if } wR_a v \text{ then } \mathscr{M}, v \vDash_\mu \varphi; \\
\mathscr{M}, w \vDash_\mu \boxed{d}\varphi & \text{iff} & \text{for all } v \in W, \text{ if } wR_d v \text{ then } \mathscr{M}, v \vDash_\mu \varphi.
\end{array}$$

For a complete axiomatization of FOIL, we refer the reader to Fitting 2006, §5.[16]

Despite having only a single domain D of objects, we can capture the idea of world-relative varying domains with the standard device of a one-place *existence predicate* E, thinking of $V(\mathsf{E}, w)$ as the non-empty domain of objects that exist in world w. We can then define the *actualist quantifier* by $\forall_a x\varphi := \forall x(\mathsf{E}x \to \varphi)$.[17]

Instead of using *role variables* $r_1, r_2, \dots$, a number of authors use what we call *role-based object variables* $z_1, z_2, \dots$.[18] These are treated in the same way by an assignment μ, so that $\mu(z_i) = \mu(r_i) \in F$, but their interpretations differ:

$$\begin{aligned}
[r_i]_{\mathscr{M},w,\mu} &= \mu(r_i); \\
[z_i]_{\mathscr{M},w,\mu} &= \mu(z_i)(w).
\end{aligned}$$

Hence r_i picks out a role, whereas z_i picks out the object that plays that role, which may vary between worlds. The truth clause for quantification with z variables is:

$$\mathscr{M}, w \vDash_\mu \forall z_i \varphi \ \text{ iff } \ \text{for all } f \in F,\ \mathscr{M}, w \vDash_{\mu[z_i/f]} \varphi.$$

---

[16] Some minor changes must be made, e.g., since we include constants in the language (recall note 13), but we will not go into the details here.

[17] One may then wish to add the assumption that for all $f \in F$, if $f(w) = d$, then $d \in V(\mathsf{E}, w)$, i.e., if an object d plays a role for the agent in w, then d exists in w, validating $\mathsf{P}(t, r_i) \to \exists_a x\, t = x$.

[18] See, for example, Parks 1974, Carlson 1988, Priest 2002, §8, and Aloni 2005.

One complication with this approach is that if we allow F to contain *partial* functions, then $\mu(z_i)$ may be undefined at w, in which case $[z_i]_{\mathscr{M},w,\mu}$ is undefined. Hence we have to deal with evaluating atomic formulas containing undefined terms. Typically authors who use such variables assume that F is a set of total functions.[19]

Let us see how we can translate a language with z variables into our language.

Definition 6 (Z-translation and abbreviation). For an atomic formula $At(t_1,\ldots,t_n)$, possibly containing x variables and z variables (instead of r variables), let $\mathbf{z}$ be the sequence of z variables contained therein (in order of first occurrence); let I be the set of indices of these variables; and let $\mathbf{x}$ be the sequence of x variables obtained from $\mathbf{z}$ by replacing each z_i with $x_{i^\star}$, where $i^\star$ is the least $j \geq i$ such that x_j is not in $At(t_1,\ldots,t_n)$ and $j \neq k^\star$ for any z_k preceding z_i in $\mathbf{z}$. Define the *Z-translation*:

$$Z(At(t_1,\ldots,t_n)) = \exists\mathbf{x}(\bigwedge_{i\in I}\mathsf{P}(x_{i^\star},r_i) \wedge At(t_1,\ldots,t_n)^{\mathbf{z}}_{\mathbf{x}}),$$

where $\exists\mathbf{x}$ abbreviates a string of quantifiers and $\varphi^{\mathbf{z}}_{\mathbf{x}}$ indicates simultaneous substitution of each element of $\mathbf{x}$ for the corresponding element of $\mathbf{z}$, and

$$\begin{aligned}
Z(\neg\varphi) &= \neg Z(\varphi);\\
Z((\varphi \wedge \psi)) &= (Z(\varphi) \wedge Z(\psi));\\
Z(\forall x_i \varphi) &= \forall x_i Z(\varphi);\\
Z(\forall z_i \varphi) &= \forall r_i Z(\varphi);\\
Z(\Box\varphi) &= \Box Z(\varphi).
\end{aligned}$$

If $\alpha = Z(\beta)$, then we call β a *Z-abbreviation* of α.

For example, our formula from §2.3,

(26) $\mathit{Tully} = z_1 \wedge \boxed{d}\mathit{Rom}(z_1) \wedge \mathit{Cicero} = z_2 \wedge \neg\boxed{d}\mathit{Rom}(z_2)$,

is the Z-abbreviation of

(27) $\exists x_1(\mathsf{P}(x_1,r_1) \wedge \mathit{Tully} = x_1) \wedge \boxed{d}\exists x_1(\mathsf{P}(x_1,r_1) \wedge \mathit{Rom}(x_1)) \wedge$
$\exists x_2(\mathsf{P}(x_2,r_2) \wedge \mathit{Cicero} = x_2) \wedge \neg\boxed{d}\exists x_2(\mathsf{P}(x_2,r_2) \wedge \mathit{Rom}(x_2))$,

which is equivalent to

(28) $\mathsf{P}(\mathit{Tully},r_1) \wedge \boxed{d}\exists x_1(\mathsf{P}(x_1,r_1) \wedge \mathit{Rom}(x_1)) \wedge$
$\mathsf{P}(\mathit{Cicero},r_2) \wedge \neg\boxed{d}\exists x_2(\mathsf{P}(x_2,r_2) \wedge \mathit{Rom}(x_2))$.

We will use Z-abbreviations repeatedly in order to compare our formalizations of belief ascriptions with those of other authors whose systems use z variables.[20]

---

[19] The exception among the authors referenced in note 18 is Carlson, who allows F to contain partial functions and uses a three-valued semantics to deal with undefined terms.

[20] Remember that in (27) and (28), we read $\boxed{d}\varphi$ as "it is doxastically necessary that φ" or "in all worlds compatible with the agent's beliefs, φ." The whole of (27) gives the condition that the truth of the belief report imposes on the actual world and the space of the agent's doxastic alternatives, so we would not read the second conjunct as "the agent believes that there exists . . . ."

There has been much discussion in the literature on quantified modal logic about whether quantifiers should take regular object variables x_i or what we called role-based object variables z_i (see Garson 2001, 2006, §13). We think both types of quantification are useful.[21] In §5, we will suggest that trying to make quantification with z variables also do the normal work of quantification with x_i variables forces one to postulate otherwise unnatural constraints on the functions assigned to z_i variables.

There is another important point about variables and quantifiers that will apply throughout. For simplicity, the belief reports we consider are cases of *role provision*,[22] where the relevant roles are clear from context, so we formalize the report with an *open* formula with *free* role variables to which the relevant roles are assigned by a context-sensitive assignment μ. In other cases of *role constraint*, there may not be any unique roles clear from context, although context supplies a set of relevant roles, so we formalize the report by prefixing role quantifiers; e.g., (28) goes to

(29) $\exists r_1 \mathsf{P}(\mathit{Tully}, r_1) \wedge \boxed{d}\exists x_1(\mathsf{P}(x_1, r_1) \wedge \mathit{Rom}(x_1)) \wedge$
$\exists r_2 \mathsf{P}(\mathit{Cicero}, r_2) \wedge \neg\boxed{d}\exists x_2(\mathsf{P}(x_2, r_2) \wedge \mathit{Rom}(x_2))$.

Then we can think of the set F of roles over which the quantifiers range as context sensitive, so that a change of context can be represented by a change of models from $\mathscr{M} = \langle W, R_a, R_d, D, F, V\rangle$ to $\mathscr{M} = \langle W, R_a, R_d, D, F', V\rangle$ as in dynamic epistemic logic. Or we could put many sets of roles in one model and superscript the role variables so that, e.g., r_k^i and r_k^j range over sets $\mu(i)$ and $\mu(j)$ of roles, where the context-sensitive assignment μ also maps numbers (superscripts) to sets of roles, as in Aloni's analogous approach with z-variables at the end of §5. Then the quantifiers in a sentence could range over distinct role sets. Role provision could be seen as the special case of role constraint where the cardinality of the relevant set of roles is 1.

3.1 Extension for Counterfactual Attitudes

The framework presented so far is designed to handle doxastic/epistemic attitudes. As noted in §2.3, Ninan (2012) has raised a challenge for accounts of these attitudes in the style of Kaplan and Lewis, on the grounds that they do not generalize to handle "counterfactual attitudes" like *imagining*, *wishing*, and *dreaming*.

To see the problem, let us return to the example of Ralph and Ortcutt from §2.3. Suppose Ralph looks at Ortcutt on the beach and imagines that *no one ever saw that man*. We might report this as

(30) Ralph imagines that no one ever saw Ortcutt.

---

[21] Belardinelli and Lomuscio (2009) include both x and z variables in their multi-agent quantified epistemic logic. Instead of distinguishing two types of variables, we could instead distinguish two types of quantifiers, in the tradition of Hintikka's (1969) distinction between $\exists y$ and Ey. By understanding $\exists z$ quantification in terms of agent-relative roles, we are following Perry (2009).

[22] This point is inspired by Crimmins and Perry (1989) on notion provision vs. notion contraint.

Ninan suggests that the truth-conditions of (30) should be stated in terms of what is true in the *worlds compatible with what Ralph is imagining*. By analogy with our treatment of the Tully-Cicero case in (28), we might try to formalize (30) as (31), relative to an assignment μ such that $\mu(r)$ is the role of *being the man seen*:

(31) $\mathsf{P}(Ortcutt, r) \wedge \boxed{i}\exists x(\mathsf{P}(x,r) \wedge NeverSeen(x))$,

where $\boxed{i}\varphi$ means that φ holds in all worlds compatible with Ralph's imagining. But (31) won't work, because it should be false throughout our intended model; since $\mu(r)$ is supposed to be the role of *being the man seen*, no one who plays that role in a world can be in the extension of the *NeverSeen* predicate in that world.

The reason that the schema for belief, represented in (28), does not work for imagination is that the schema for belief commits us to the following.

> **Doxastic Match Hypothesis**: for an agent to believe of an object *o* via a role **r** that it has a property, the agent must also believe of *o* via **r** *that it plays* **r**.

As a corollary, if the agent does not have *any* true beliefs about the roles that *o* plays in his life, then the agent cannot have any *de re* beliefs about *o*. We think this hypothesis is plausible for a suitably general notion of "belief" (positive doxastic attitude), but Ninan is right to reject the analogous hypothesis for imagination.

> **Imaginative Match Hypothesis**: for an agent to "imagine of" object *o* via role **r** that it has a property, the agent must also imagine of *o* via **r** *that it plays* **r**.

The example above, where Ralph sees Ortcutt and imagines that no one ever saw that man, is a counterexample to the Imaginative Match Hypothesis. It is not clear, however, whether Ralph can see Ortcutt and *believe* that no one ever saw that man.

Inspired by Lewis (1983), Ninan's solution to the problem of counterfactual attitudes is (roughly) the following: we can *stipulate* that an object in a given "imagination world" *v* is supposed to represent Ortcutt, rather than finding the object as the player in *v* of a role that Ortcutt plays in the actual world. On this view, to check at a world whether Ralph is imagining that Ortcutt has a given property, we simply check whether the *stipulated representatives* of Ortcutt across Ralph's imagination worlds have that property. But there is an additional subtlety, arising due to "double vision" cases. Recall that for *Ralph*, Ortcutt plays both the roles of *being the man seen* and *being the man called 'Bernard Ortcutt'*. Suppose Ralph looks at Ortcutt on the beach and says: "I'm imagining a situation in which that guy [Ralph points at the man on the waterfront] is distinct from Bernard Ortcutt, and in which I never saw that guy, and in which Ortcutt never goes by the name 'Bernard Ortcutt"' (Ninan, 2012, 24). For handling this kind of case, Ninan generalizes the proposal stated above: we can stipulate that one object in an imagination world *v* represents Ortcutt relative to one role (*being the man seen*), while a different object in *v* can represent Ortcutt relative to another role (*being the man called 'Bernard Ortcutt'*).

Let us sketch how Ninan's treatment of counterfactual attitudes can be accommodated in the framework of FOIL. For simplicity, in §3 we did not define the full language of FOIL as presented by Fitting. In the full language, there are *intensional predicates* that we can apply to function variables. Previously we called function variables 'role variables', but let us now consider adding partial functions to F in our models that are not thought of as roles. We introduce a two-place intensional predicate Stip such that for function variables r and s,[23] $\mathsf{Stip}(r,s)$ is a formula. Intuitively, think of the truth of $\mathsf{Stip}(r,s)$ at world w relative to an assignment μ as telling us the following about the function $\mu(s)$: in every world v, $\mu(s)(v)$ is the individual in v that "represents" $\mu(r)(w)$, relative to role $\mu(r)$, "by stipulation."[24] In essence, Ninan proposes truth-conditions for (30) that are equivalent to that of

(32) $\mathsf{P}(Ortcutt,r) \wedge \exists s(\mathsf{Stip}(r,s) \wedge \boxed{i}\exists x(\mathsf{P}(x,s) \wedge NeverSeen(x)))$.

In other words, Ortcutt plays the role of *being the man seen* for Ralph, and in every world compatible with what Ralph imagines, the thing that *by stipulation* represents Ortcutt *relative to that role* has the property of never being seen. It is easy to see how this approach will also handle Ralph's double vision case above.

In what follows, we will return to the simpler treatment of belief that assumes the Doxastic Match Hypothesis. However, it is noteworthy that the FOIL framework has the flexibility to model attitudes for which such a hypothesis is not reasonable.

4 Names in Alethic and Epistemic Logic

In this section, we discuss the treatment of names in alethic and epistemic/doxastic logic. In §4.1, we begin by discussing a problem about names that has lead some to reject standard semantics for epistemic/doxastic predicate logic as incoherent. In §4.4, we will propose a solution to this problem based on the ideas of §2.3.

4.1 The "Hintikka-Kripke Problem"

Consider the difference between the following:

(33) Hesperus = Phosphorus, but it's not necessary that Hesperus = Phosphorus.
(34) Hesperus = Phosphorus, but Elwood doesn't believe that Hesperus = Phosphorus.

Following Kripke (1980), (33) should be formalized by a sentence in our formal language that is *unsatisfiable*. By contrast, following Hintikka (1962), (34) should be formalized by a sentence that is *satisfiable*. Translating (33) as

(35) $Hesperus = Phosphorus \wedge \neg\boxed{a}Hesperus = Phosphorus$,

[23] We could have two sorts of function variables, $r_1, r_2, \dots$ for roles and $s_1, s_2, \dots$ for non-role functions. Or we could indicate the difference between role functions and non-role functions by a one-place predicate Role whose extension contains only functions to be thought of as roles.

[24] According to this intuitive understanding, the extension of Stip should be a *functional* relation: if $\mathsf{Stip}(r,s)$ holds, then $\mathsf{Stip}(r,s')$ should not hold for any $s' \neq s$.

if we assume the R_a*-rigidity of names* from Definition 2, then (35) is unsatisfiable, as desired. Similarly, if we translate (34) as

(36) $Hesperus = Phosphorus \wedge \neg \boxed{d} Hesperus = Phosphorus$

and assume the R_d*-rigidity of names*, then (36) is also unsatisfiable—but we want our formalization of (34) to be *satisfiable*. This has been called the "Hintikka-Kripke Problem" (Lehmann, 1978). To solve it, we have two choices: give a different formalization of (34) or give up R_d-rigidity. The second seems to have become common among logicians working on epistemic predicate logic, following Hintikka's (1970, 872) view that "in the context of propositional attitudes even grammatical proper names do not behave like 'logically proper names'." For example, in their textbook on first-order modal logic, Fitting and Mendelsohn (1998, 219) write that "the problem, quite clearly, lies with the understanding that these names are rigid designators. We see that, although they are rigid within the context of an *alethic* reading of $\Box$, they cannot be rigid under an *epistemic* reading of $\Box$." This suggests that we should not assume the R-rigidity of names when R is an epistemic/doxastic accessibility relation (also see Priest 2002, 456f). The question is how we are supposed to understand the failure of R_d-rigidity. As Linksy (1979, 92) writes:

> Hence, these names do not denote the same thing in all doxastically possible worlds; that is, they are not doxastically rigid designators. How are we to make this situation intelligible to ourselves? Hesperus (= Phosphorus) is (are?) *two* objects in the world described by the sentences [true at some world]. It is not just that 'Hesperus' and 'Phosphorus' are names of different objects, for that is easily enough understood. The problem is that in this world Hesperus (= Phosphorus) is not Phosphorus (= Hesperus). That cannot be understood at all.

Let us call this the *impossible worlds* explanation of the failure of R_d-rigidity, which leads Linsky to believe that there are problems with analyzing belief ascriptions in terms of the semantics of $\Box$ that do not arise for alethic necessity (also see Barnes 1976, which is even more critical). By contrast, Aloni (2005, 9-10) interprets the failure of R_d-rigidity in the weaker way that Linsky mentions:

> [I]n different doxastic alternatives a proper name can denote different individuals. The failure of substitutivity of co-referential terms (in particular proper names) in belief contexts does not depend on the ways in which terms actually refer to objects (so this analysis is not in opposition to Kripke's (1972) theory of proper names), *it is simply due to the possibility that two terms that actually refer to one and the same individual are not believed by someone to do so* .... Many authors ... have distinguished semantically rigid designators from epistemically rigid designators—the former refer to specific individuals in counterfactual situations, the latter identify objects across possibilities in belief states—and concluded that proper names are rigid only in the first sense.[25] [emphasis added]

The emphasized claim is in the tradition of Stalnaker (1984, 85f): "If a person is ignorant of the fact that Hesperus is Phosphorus, it is because his knowledge fails to exclude a possible situation in which, because causal connections between names and objects are different, one of those names refers to a different planet, and so the statement, 'Hesperus is Phosphorus' says something different than it actually says."

[25] The last of the quoted sentences occurs in footnote 7 of Aloni 2005.

However, there are two problems with appealing to Stalnaker's idea in support of making (36) satisfiable by giving up R_d-rigidity: first, the appeal leads to serious problems about how to understand the valuation function V in modal models and about how R_d and R_a are related, as discussed in §4.2; second, Stalnaker's idea is inadequate to explain other attributions of ignorance, as discussed in §4.3. After explaining these problems, we will solve the Hintikka-Kripke problem in the way suggested by our discussion in §2.3: give a different formalization of (34).

4.2 Representational vs. Interpretational Semantics

Those who give up R_d-rigidity allow models in which, e.g., wR_dv, $V(Hesperus,w) = V(Phosphorus,w)$, but $V(Hesperus,v) \neq V(Phosphorus,v)$. One might think that this simply reflects Stalnaker's idea that 'Hesperus' and 'Phosphorus' refer to different things in v but the same thing in w. The problem is that this blurs an important distinction about how to understand the valuation V. This distinction is closely related to Etchemendy's (1999) distinction between *representational* and *interpretational* semantics for classical logic, but here applied to models for modal logic.

When we switch from one model $\mathscr{M} = \langle W,R,D,F,V\rangle$ to another model $\mathscr{M}' = \langle W,R,D,F,V'\rangle$ that differs only with respect to its valuation function, one way to think of this switch is as a change in how we interpret the language. Intuitively, V may interpret the predicate *white* so that its extension $V(white,w)$ in world w is the set of things in w that are white, while V' may interpret the predicate *white* so that its extension $V'(white,w)$ in world w is the set of things in w that are green.

When we switch from one world w to another world v within the same model $\mathscr{M} = \langle W,R,D,F,V\rangle$ (so $w,v \in W$), the extension $V(white,w)$ of the predicate *white* in w may differ from the extension $V(white,v)$ of *white* in v. If this is so, then the set of white objects in w differs from the set of white objects in v. Similarly, if $V'(white,w)$ differs from $V'(white,v)$, then the set of green objects in w differs from the set of green objects in v. We could also have $V(white,w) \neq V(white,v)$ but $V'(white,w) = V'(white,v)$, in which case w and v differ with respect to which objects are white, but they are the same with respect to which objects are green.

The point is that we think of the difference between $V(white,w)$ and $V(white,v)$ as reflecting a difference in how the worlds w and v are, not a difference in how we interpret the language—whereas we can think of the difference between $V(white,w)$ and $V'(white,w)$ as reflecting a difference in how we interpret the language.

Exactly the same points apply to names. If we think of the difference between $V(white,w)$ and $V(white,v)$ as reflecting a difference in how the worlds w and v are, not a difference in how we interpret the language, then we should also think of any difference between $V(Hesperus,w)$ and $V(Hesperus,v)$ as reflecting a difference in how the worlds w and v are, not a difference in how we interpret the language. (Though of course we can think of $V(Hesperus,w) \neq V'(Hesperus,w)$ as reflecting a difference in how we interpret the language.) This brings us back to Linsky's question of how we are to make intelligible that $V(Hesperus,w) = V(Phosphorus,w)$

but $V(Hesperus, v) \neq V(Phosphorus, v)$. For in this case v is not merely a world in which words refer to different things; it is a metaphysically impossible world.

Here is a crucial point: assuming that $V(Hesperus, w) = V(Phosphorus, w)$ but $V(Hesperus, v) \neq V(Phosphorus, v)$, we cannot have $wR_a v$, for otherwise we violate the R_a-rigidity of names and (35) becomes satisfiable, which no one wants. Hence it is not the case that $wR_a v$, so v is an impossible world (relative to w). This shows that v cannot simply be understood as a world largely like w but in which 'Hesperus' refers to something different than in w—for such a world should be metaphysically *possible* (relative to w). The upshot is that one cannot explain the failure of R_d-rigidity in the weaker way proposed at the end of §4.1.

A defender of that proposal might reply that the problem is instead with trying to combine alethic and doxastic operators in one system, as we have. But why should such a combination be problematic? We take it to be a virtue of our approach below that it allows such a combination for a full solution of the Hintikka-Kripke problem.

Although Stalnaker's idea from §4.1 cannot be implemented in standard modal semantics simply by giving up R_d-rigidity, it may be possible to implement with a more complicated two-dimensional framework. However, we will not investigate that possibility until §4.5, since in §4.3 we will raise a problem for the idea itself.

4.3 Ignorance of Co-Reference vs. Ignorance of Identity

The problem is that ignorance of co-reference and ignorance of identity can come apart. Before explaining this idea, let us formally represent ignorance of co-reference. Suppose that our language contains not only names $n_1, n_2, \ldots$ of objects, but also names 'n_1', 'n_2', ... of those names. Moreover, suppose that the object names $n_1, n_2, \ldots$ are also elements of the domain D of our models—they are also objects—and let us require that $V(\text{`}n_i\text{'}, w) = n_i$ for all names n_i and worlds w. Finally, suppose we have a two place naming predicate N such that $V(N, w)$ is the set of pairs $\langle n_i, d\rangle$ such that speakers in world w use n_i to *name* object d. With this setup, we can state the point from §4.2 about how to understand the V function as follows:

$\langle n_i, d\rangle \in V(N, w)$ need not imply $V(n_i, w) = d$ or vice versa.

In other words, how speakers in different worlds in our model use names in the language is one thing, reflected by $V(N, w)$; and how *we* who build the model are interpreting the names of the language is another, reflected by $V(n_i, w)$.

Now we can easily represent a world w in which an agent does not believe that 'Hesperus' and 'Phosphorus' co-refer by including a world v such that $wR_d v$ and:

- $\langle Hesperus, d\rangle \in V(N, w)$ and $\langle Phosphorus, d\rangle \in V(N, w)$;
- $\langle Hesperus, d\rangle \notin V(N, v)$ and $\langle Phosphorus, d\rangle \in V(N, v)$.

In such a case, even given that $V(Hesperus, u) = V(Phosphorus, u)$ for all $u \in W$, the following are both satisfied at w:

(37) $\forall x(N(\text{`}Hesperus\text{'}, x) \leftrightarrow N(\text{`}Phosphorus\text{'}, x))$;

(38) $\neg \boxed{d} \forall x(N(\text{‘}Hesperus\text{’},x) \leftrightarrow N(\text{‘}Phosphorus\text{’},x))$,

so we have a case of ignorance of co-reference.

Before explaining how to formalize ignorance of identity, we will show how ignorance of co-reference and ignorance of identity can come apart.

Example 1 (The Registrar). The county registrar goes fishing regularly with his old friend Elwood. Unknown to the registrar, however, Elwood's identical twin, Egbert, occasionally substitutes for him on these fishing trips. Since the trips are mostly silent, Egbert has no problem keeping the deception from the registrar. The registrar regularly says things like "Hey Elwood, pass me a beer," while talking to Egbert.

It seems fair to say,

(39) Although Egbert is not Elwood, the registrar believes that Egbert is Elwood.

Traditionally, (39) would be rendered as

(40) $Egbert \neq Elwood \wedge \boxed{d}\, Egbert = Elwood$.

The truth of (40) requires that $Egbert = Elwood$ be true in all worlds compatible with what the Registrar believes. Its truth therefore requires that names are not "doxastically rigid" designators. At first blush, this seems to mean that only impossible worlds are compatible with what the registrar believes. In the spirit of Aloni, we might seek to avoid this unfortunate result by adopting the following analysis.

> **Co-Reference Mistake Analysis**: the satisfiability of (40) is simply due to the possibility that two terms that actually refer to different individuals are believed by the registrar to co-refer, that is, to refer to the same individual.

The problem with this analysis becomes clear when we consider the rest of the story.

Example 1 (The Registrar Continued). The registrar, being the registrar, knows that Elwood has a brother, whom he thinks he has never met, although he sends him an invoice for his taxes each year, and often follow-up reminders. Based on his records, the registrar *knows* that 'Elwood' and 'Egbert' refer to different people.

Hearing the registrar say to Egbert, "Hey Elwood, your brother better pay his taxes," someone in on the deception might explain to a third party:

(41) Although the registrar knows that 'Elwood' and 'Egbert' refer to different people, he believes that Egbert is Elwood.

What (41) shows is that the Co-Reference Mistake Analysis doesn't work. The truth of (39) cannot always be understood in terms of false belief about co-reference, because the truth of (41) cannot be so understood. In §4.4, we will explain how, on our account, (41) can nonetheless be a reasonable and true thing to say.

4.4 The "Hintikka-Kripke Problem" Resolved

In §4.1, we noted two ways to respond to the Hintikka-Kripke Problem: give up R_d-rigidity or give a different formalization of (34). Having seen the problems with the first, let us consider the second. When someone says "Elwood does not believe that Hesperus is Phosphorus," our formalization of the claim is not

(42) $\neg \boxed{d}\, Hesperus = Phosphorus.$

In fact, it is not clear that (42) is the correct formalization of any natural language belief ascription. As Lewis (1977, 360) said in another context, "why must every logical form find an expression in ordinary language?" Relatedly, consider:

(43) $Hesperus = Hesperus \to \boxed{d}\, Hesperus = Hesperus.$
(44) $Hesperus = Phosphorus \to \boxed{d}\, Hesperus = Phosphorus.$

In our framework, (43) *and* (44) are valid, but this does not mean that we would claim in natural language that "If Hesperus is Phosphorus, then Elwood believes that Hesperus is Phosphorus." For the function of belief ascriptions involving names in natural language is not just to say something about how the *objects named* show up across doxastic alternatives, which is all that (44) manages to capture.

Before handling the Hesperus and Phosphorus case, let us treat the story of Elwood and Egbert from Example 1, using ideas from §2.3. Recall:

(39) Although Egbert is not Elwood, the registrar believes that Egbert is Elwood.

In the context of Example 1, we propose to analyze (39) as follows. Elwood is playing a number of epistemic roles in the registrar's life, but the role that is contextually salient in Example 1 is *being the source of a cluster of memories—about his old friend and fishing buddy* ($\mathbf{r}_f$). Egbert is also playing a number of epistemic roles in the registrar's life, but the role that is contextually salient in Example 1 is *being the person seen and talked to* ($\mathbf{r}_s$). Consider a model $\mathscr{M}$, assignment μ, and role variables r_f and r_s such that for all worlds w in $\mathscr{M}$, $\mu(r_f)(w)$ is the player of role $\mathbf{r}_f$ in w, and $\mu(r_s)(w)$ is the player of role $\mathbf{r}_s$ in w. Then using Z-abbreviation (Definition 6), we formalize (39) in the style of (26) in §2.3, instead of (40):

(45) $Egbert \neq Elwood \wedge Elwood = z_f \wedge Egbert = z_s \wedge \boxed{d}\, z_f = z_s.$

If we translate (39) as (45), we can maintain the Conservative Approach of §1: there is no need to claim that names in doxastic logic are not "doxastically rigid" or that "doxastically possible worlds" are not plain possible worlds. Moreover, (45) can be true even if the registrar believes that 'Elwood' and 'Egbert' do not co-refer:

(46) $\boxed{d}\, \neg\exists x_1 (N(\text{'Elwood'}, x_1) \wedge N(\text{'Egbert'}, x_1)).$

Hence we can also handle the second part of the registrar story in (41).

Let us now return to the classic case of Hesperus and Phosphorus:

(47) Elwood does not believe that Hesperus is Phosphorus.

One might utter (47) in a variety of contexts. Suppose Elwood has never heard the words ‘Hersperus’ and ‘Phosphorus’, but he likes to look at the planet Venus early in the evening and at the same planet early in the morning, not realizing it is the same one. In this case, it would be natural to utter (47). Take a model $\mathscr{M}$, assignment μ, and role variables r_e and r_m such that for any world w in $\mathscr{M}$, $\mu(r_e)(w)$ is the star that Elwood likes to look at in the evening in w, and $\mu(r_m)(w)$ is the star that Elwood likes to look at in the morning in w. Using Z-abbreviation, we translate (47) as:

(48) $Hesperus = z_e \wedge Phosphorus = z_m \wedge \neg \boxed{d}\, z_e = z_m.$

Another context in which (47) makes sense is one where Elwood has heard of Hesperus and Phosphorus in Astronomy class, but since he wasn’t paying attention, he has no idea what they are. Here the roles that Hesperus/Phosphorus plays in his life are simply *being the source of the ‘Hesperus’ network exploited by Elwood* (recall §2.3) and *being the source of the ‘Phosphorus’ network exploited by Elwood*. In this case, we translate (47) with a sentence of the same form as (48), only using role variables associated with these different roles in our model. What this shows is that the role-based analysis subsumes Stalnaker’s (1984, 85f) analysis (recall §4) in those contexts where Stalnaker’s analysis works. But the role-based analysis also works in cases like the Elwood-Egbert story, where Stalnaker’s does not.

Finally, since (48) can be true at a world where

(49) $\boxed{a}\, Hesperus = Phosphorus$

is true, belief attributions as in (47) do not pose a problem for a combined epistemic-alethic modal logic. The Hintikka-Kripke Problem is no longer a problem.

4.5 Two-Dimensional Epistemic Models

Recall Aloni’s (2005, 9) idea, similar to Stalnaker’s (1984, 85f), that Hesperus-Phosphorus style cases arise because of the “possibility that two terms that actually refer to one and the same individual are not believed by someone to do so.” In §4.2, we argued that this idea cannnot be correctly implemented in standard modal semantics. Let us now return to the suggestion that it may be implementable in a two-dimensional epistemic framework (cf. Schroeter 2012 and references therein).

Definition 7 (2D Constant Domain Models). A 2D (constant domain) model for the language of Definition 1 is a tuple $\mathscr{M} = \langle W, R_a, \mathbf{R}_d, D, F, \mathbf{V}\rangle$ where W, R_a, D, and F are defined as in Definition 2; $\mathbf{R}_d$ is a binary relation on $W \times W$; for any relation symbol Q_i^k and worlds $w, v \in W$, $\mathbf{V}(Q_i^k, w, v)$ is a set of k-tuples of objects; for any name n_i and worlds $w, v \in W$, $\mathbf{V}(n_i, w, v)$ is an object. We assume the general rigidity condition for names that for all $w, v, u \in W$, $\mathbf{V}(n_i, w, v) = \mathbf{V}(n_i, w, u)$.

For the sake of generality, we have defined $\mathbf{R}_d$ as a kind of 2D accessibility relation, following Israel and Perry (1996) and Rabinowicz and Segerberg (1994). This raises the question of what $\langle w, v\rangle \mathbf{R}_d \langle w', v'\rangle$ is supposed to mean intuitively.

However, here we will consider the class of models such that if $\langle w,v\rangle \mathbf{R}_d \langle w',v'\rangle$, then $w' = v'$ and $\langle x,v\rangle \mathbf{R}_d \langle w',v'\rangle$ for all $x \in W$. Hence all we need to know is whether the second coordinates are related, written as $v\mathbf{R}_d v'$. Take $v\mathbf{R}_d v'$ to mean that everything the agent believes in v is compatible with the hypothesis that v' is his actual world.

For a name n_i, take $\mathbf{V}(n_i,w,v) = d$ to mean that if we consider w as the actual world, then n_i (as used by speakers in w) names d in world v. Suppose w is a world in which the heavenly body that people see in the evening, that they call 'Hesperus', etc., is the same as the heavenly body that they see in the morning, that they call 'Phosphorus', etc. Hence if we consider w as actual, then we will have $\mathbf{V}(Hesperus,w,v) = \mathbf{V}(Phosphorus,w,v)$ for all worlds $v \in W$, regardless of how language is used (or whether there are any language users) in v. However, if we consider as actual a world w' in which the heavenly body that people see in the morning, that they call 'Hesperus', etc., is not the same as the heavenly body that they see in the morning, that they call 'Phosphorus', etc., then we will have $\mathbf{V}(Hesperus,w',v) \neq \mathbf{V}(Phosphorus,w',v)$ for all worlds $v \in W$.[26]

Variable assignments are defined as in Definition 3, but the interpretation of a name is now given relative to a pair of worlds instead of a single world.

Definition 8 (Interpretation of Terms). The interpretation $[t]_{\mathscr{M},w,v,\mu}$ of a term t in a 2D model $\mathscr{M}$ at world v, *with world w considered as actual*, is an object given by:

$$[n_i]_{\mathscr{M},w,v,\mu} = V(n_i,w,v);$$
$$[x_i]_{\mathscr{M},w,v,\mu} = \mu(x_i).$$

Definition 9 (2D Truth). Given a 2D model $\mathscr{M}$ with $w,v \in W$, an assignment μ, and a formula φ, we define $\mathscr{M},w,v \vDash_\mu \varphi$ as follows:

$$
\begin{array}{lll}
\mathscr{M},w,v \vDash_\mu Q_i^k(t_1,\ldots,t_k) & \text{iff} & \langle [t_1]_{\mathscr{M},w,v,\mu},\ldots,[t_k]_{\mathscr{M},w,v,\mu}\rangle \in V(Q_i^k,w,v);\\
\mathscr{M},w,v \vDash_\mu t = t' & \text{iff} & [t]_{\mathscr{M},w,v,\mu} = [t']_{\mathscr{M},w,v,\mu};\\
\mathscr{M},w,v \vDash_\mu \mathsf{P}(t,r_i) & \text{iff} & \mu(r_i) \text{ is defined at } v \text{ and } [t]_{\mathscr{M},w,v,\mu} = \mu(r_i)(v);\\
\mathscr{M},w,v \vDash_\mu \neg\varphi & \text{iff} & \mathscr{M},w,v \nvDash_\mu \varphi;\\
\mathscr{M},w,v \vDash_\mu (\varphi \wedge \psi) & \text{iff} & \mathscr{M},w,v \vDash_\mu \varphi \text{ and } \mathscr{M},w,v \vDash_\mu \psi;\\
\mathscr{M},w,v \vDash_\mu \forall x_i\varphi & \text{iff} & \text{for all } d \in D,\ \mathscr{M},w,v \vDash_{\mu[x_i/d]} \varphi;\\
\mathscr{M},w,v \vDash_\mu \forall r_i\varphi & \text{iff} & \text{for all } f \in F,\ \mathscr{M},w,v \vDash_{\mu[r_i/f]} \varphi;\\
\mathscr{M},w,v \vDash_\mu \boxed{a}\varphi & \text{iff} & \text{for all } v' \in W, \text{ if } vR_av' \text{ then } \mathscr{M},w,v' \vDash_\mu \varphi;\\
\mathscr{M},w,v \vDash_\mu \boxed{d}\varphi & \text{iff} & \text{for all } w',v' \in W, \text{ if } \langle w,v\rangle \mathbf{R}_d \langle w',v'\rangle\\
 & & \qquad \text{then } \mathscr{M},w',v' \vDash_\mu \varphi.
\end{array}
$$

Given our assumed constraints on $\mathbf{R}_d$, we can re-write the last clause as[27]

[26] Similarly, consider one-place predicates *ContainsWater* and *ContainsH*20. If we consider as actual a world w in which the substance that fills the lakes and rivers, is called 'water', etc., is H20, then we will have $\mathbf{V}(ContainsWater,w,v) = \mathbf{V}(ContainsH20,w,v)$ for all $v \in W$. However, if we consider as actual a world w' in which the substance that fills the lakes and rivers, is called 'water', etc., is XYZ, then we may have $v \in W$ such that $\mathbf{V}(ContainsWater,w',v) \neq \mathbf{V}(ContainsH20,w',v)$.

[27] Compare this to the "fixedly actually" operator of Davis and Humberstone (1980).

$$\mathscr{M}, w, v \vDash_\mu \boxed{d}\varphi \ \text{ iff } \ \text{for all } v' \in W, \text{ if } v\mathbf{R}_d v' \text{ then } \mathscr{M}, v', v' \vDash_\mu \varphi.$$

To see how this is supposed to solve the Hintikka-Kripke problem, consider

(50) $\boxed{a}n_1 = n_2 \wedge \boxed{d}n_1 \neq n_2$.

We have $\mathscr{M}, w, w \vDash_\mu \boxed{a}n_1 = n_2 \wedge \boxed{d}n_1 \neq n_2$ iff both of the following hold:

- for all $x \in W$, if $wR_a x$ then $\mathscr{M}, w, x \vDash_\mu n_1 = n_2$;
- for all $y \in W$, if $w\mathbf{R}_d y$ then $\mathscr{M}, y, y \vDash_\mu n_1 \neq n_2$.

Clearly we can construct a model satisfying these conditions, so (50) is satisfiable.

There is much to be said about the 2D approach, but we will limit ourselves to two points. First, there is a way of understanding the 2D treatment of names in epistemic contexts as a *special case* of the role-based treatment. In particular, with every name n_i we can associate a role f_i such that

$$f_i(v) = \mathbf{V}(n_i, v, v).$$

Then if we map role variables r_1 and r_2 to f_1 and f_2, respectively, (50) will have the same truth value at any pair of worlds as its role-based translation:

(51) $\boxed{a}n_1 = n_2 \wedge \mathsf{P}(n_1, r_1) \wedge \mathsf{P}(n_2, r_2) \wedge \boxed{d}\exists x_1 \exists x_2 (\mathsf{P}(x_1, r_1) \wedge \mathsf{P}(x_2, r_2) \wedge x_1 \neq x_2)$.

Note, however, that there may be many other roles that the objects named by n_1 and n_2 play, besides f_1 and f_2. In effect, the 2D framework restricts us to just those roles.

This leads to the second point: it is not clear how the 2D framework can handle the case of the registrar in Example 1 in §4.3 without bringing in roles. In our world w, the registrar believes that 'Egbert' and 'Elwood' refer to different people, so for any world v compatible with his beliefs, $Egbert \neq Elwood$ should be true at v *considered as actual*. But then $\boxed{d}Egbert \neq Elwood$ will be true at w, so if this is the two-dimensionalists' formalization of 'the registrar believes that Egbert is not Elwood', then they face the problem raised in §4.3: in the context of Example 1 it seems true to say instead that 'the registrar believes that Egbert is Elwood'.

In what follows we return to the 1D framework. There are good reasons for multi-dimensionality to deal with terms like 'now' and 'actually', but it seems that roles are still needed to deal with belief attributions involving names.[28] In §5, we shall see how roles are also useful in formalizing belief attributions involving quantification.

5 Quantification into Epistemic Contexts

Having shown how a role-based analysis of attitude ascriptions resolves the Hintikka-Kripke Problem, we will now apply the analysis to quantification into epistemic contexts, comparing it to the analyses of Carlson (1988) and Aloni (2005).

[28] To deal in the 1D framework with an agent who does not believe, e.g., that something contains water iff it contains H20, we would need to generalize the notion of role so that *properties* (understood extensionally, intensionally, or hyper-intensionally) can play roles for an agent.

First, consider the following sentence:

(52) "The police do not know who a certain person is" (Carlson, 1988, 232).

How should we translate (52) into our formal language? As suggested in §2.3, we cannot answer this question simply by looking at the sentence out of context.

In our view, there seem to be two readings of (52), which are natural in different contexts. First, suppose the police pride themselves on keeping track of everyone in the area. However, someone has slipped through the cracks: Jones, whom the police do not know anything about. In this case, it makes sense to utter (52), understood as

(53) $\exists x_1 \neg \exists z_1 \boxed{e}\, x_1 = z_1$,

the Z-abbreviation of

(54) $\exists x_1 \neg \exists r_1 \boxed{e} \exists x_2 (\mathsf{P}(x_2, r_1) \wedge x_1 = x_2)$,

where we write $\boxed{e}$ in place of $\boxed{d}$ for *epistemic* necessity.

Second, suppose Jones now plays the role for the police of *being the suspect chased*. However, for all the police know, they could be chasing Smith instead of Jones. In this case, it makes sense to utter (52), now understood as

(55) $\exists z_1 \neg \exists x_1 \boxed{e}\, x_1 = z_1$,

the Z-abbreviation of

(56) $\exists r_1 \neg \exists x_1 \boxed{e} \exists x_2 (\mathsf{P}(x_2, r_1) \wedge x_1 = x_2)$.[29]

Which translation is better depends on the context in which (52) is uttered.

Can we translate (52) with only z variables? Carlson (1988) tries to do so with

(57) $\exists z_1 \neg \exists z_2 \boxed{e}\, z_1 = z_2$,

where $\exists z_1$ and $\exists z_2$ both quantify over the same set of functions. He allows z variables to be mapped to partial functions and uses a three-valued semantics such that (57) is satisfiable in his framework. However, with only one kind of variable in that framework, there is no way to make a distinction like the one we have made between (53) and (55). Moreover, the intuitive meaning of (57) is not at all clear.

An advantage of having quantification over both objects and functions is that when one tries to do all the work with just quantification over functions, one is tempted to impose otherwise unnatural constraints on the set of functions. For example, Carlson (1988, 244f) and Aloni (2005, 25f) propose two conditions on the set F of functions, which they interpret as *individuating functions* and *individual concepts*, respectively: an *existence* condition and a *uniqueness* condition.

Definition 10 (Existence Condition). In $\mathscr{M} = \langle W, R_a, R_d, D, F, V \rangle$, F satisfies the *existence condition* iff for all $w \in W$, $d \in D$, there is some $f \in F$ with $f(w) = d$.

---

[29] Note that (55)/(56) does not require the existence of anyone who actually plays r_1. We can express a reading that requires the existence of a role-player with: $\exists z_1 \exists x_1 z_1 = x_1 \wedge \neg \exists x_2 \boxed{e}\, x_2 = z_1$.

When we think of the functions in F as agent-relative roles, the existence condition is not plausible. For it is built in to the idea of agent-relativity that an object in a world many not play any role in the cognitive life of our agent.

Let us consider Aloni's argument for the existence condition. First, consider:

(58) If the president of Russia is a spy, then there is someone who is a spy.

Translating (58) as

(59) $S(p) \to \exists z S(z)$,

Aloni (2005, 26) notes that if we do not assume the existence condition on F, then (59) is not valid, whereas the translation of (58) should be valid. Hence we should assume the existence condition. However, this is too quick, because we have two options: assume the existence condition or give a different translation of (58). In the framework of §3, there is a clear candidate for the latter:

(60) $S(p) \to \exists x S(x)$,

where $\exists x$ quantifies over D. Unlike (59), (60) is valid in our framework. We take (60) to be the appropriate translation of (58). The existence assumption seems to be an artifact of trying to make $\exists z$ do all the work of two types of quantification.

Definition 11 (Uniqueness Condition). In $\mathscr{M} = \langle W, R_a, R_d, D, F, V \rangle$, F satisfies the *uniqueness condition* iff for all $w \in W$, $f, f' \in F$, if $f \neq f'$, then $f(w) \neq f'(w)$.

When we think of functions in F as agent-relative roles, the uniqueness condition is not plausible. For it is built in to the idea of roles that there can be an object in a world that plays multiple roles in the cognitive life of our agent.

Without uniqueness, we can easily handle ascriptions of ignorance such as:

(61) "There is someone who might be two different people as far as the police know" (Carlson, 1988, 237).

Imagine, for example, that although the same person was both the thief and the getaway driver, for all the police know, different people played these roles—it was a two man job. Corresponding to (61), we have the satisfiable sentence

(62) $\exists z_1 \exists z_2 (z_1 = z_2 \wedge \neg \boxed{e}\, z_1 = z_2)$.

As one can easily check with Z-translation, (62) is true in the case where the same person plays the role of the thief and the role of the driver in the actual world, but in some world compatible with what the police know, there are two people involved.

Although the most natural way of making (62) satisfiable violates uniqueness, Carlson manages to make (62) satisfiable while requiring uniqueness. He does so by mapping z_1 and z_2 to the same partial function, which is defined at the world of evaluation but undefined at some epistemically accessible world. As Carlson (1988, 238) puts it, "[(62)] in our interpretation does not imply that ... [z_1 and z_2] ... pick out two different properly cross-identified individuals in some alternative, only that they fail to refer to one and the same individual somewhere."

The problem with Carlson's analysis is that it does not generalize to capture other cases, such as the following elaboration of the heist example above:

(63) Someone who was at the crime scene might be two people as far as the police know, but the police know that whoever was there was a gangster.

It seems that any good formalization of (63) should imply the following:

(64) $\exists z_1 \exists z_2 (z_1 = z_2 \wedge \neg \boxed{e}\, z_1 = z_2 \wedge \boxed{e}(Gz_1 \wedge Gz_2))$.

However, (64) is unsatisfiable in Carlson's system. For if z_1 and z_2 are mapped to functions that are undefined at some epistemically accessible world, as required for $z_1 = z_2 \wedge \neg \boxed{e}\, z_1 = z_2$ to be true for Carlson, then $\boxed{e}\,(Gz_1 \wedge Gz_2)$ is not true. By contrast, since we reject uniqueness, (64) is satisfiable in our framework.

Aloni observes that without uniqueness, the following are not equivalent:

(65) $\neg \exists z_1 (z_1 = Ortcutt \wedge \boxed{d} Spy(z_1))$.
(66) $\exists z_1 (z_1 = Ortcutt \wedge \neg \boxed{d} Spy(z_1))$.

Indeed, without uniqueness, the truth of (66) is compatible with the falsity of (65). And without existence, the truth of (65) is compatible with the falsity of (66).

Given the non-equivalence of (65) and (66), Aloni (2005, 24) concludes that formal systems without the uniqueness condition "predict a structural ambiguity for sentences like 'Ralph does not believe Ortcutt to be a spy', with a wide scope reading asserting that Ralph does not ascribe espionage to Ortcutt under any (suitable) representation, and a narrow scope reading asserting that there is a (suitable) representation under which Ralph does not ascribe espionage to Ortcutt. This ambiguity is automatically generated by any system" that does not satisfy uniqueness, but Aloni doubts that there is any such ambiguity in natural language.

But the fact that (65) and (66) are not equivalent in a given formal system does not mean that the system "predicts a structural ambiguity" for the English sentence

(67) Ralph does not believe Ortcutt to be a spy.

Linguists predict ambiguities in English. Logical systems do not. (Repeating Lewis (1977), "why must every logical form find an expression in ordinary language?")

On our view, the correct formalization of (67) depends on the context. Suppose we just had a long conversation about Ralph's only next door neighbor, Ortcutt, when you utter (67). Although Ortcutt plays several roles in Ralph's life, in this case the contextually salient role is *being the next door neighbor*. Consider a model $\mathscr{M}$, assignment μ, and role variable r_n such that for any world w in $\mathscr{M}$, $\mu(r_n)(w)$ is Ralph's next door neighbor in w. Using Z-abbreviation, we formalize (67) as

(68) $Ortcutt = z_n \wedge \neg \boxed{d} Spy(z_n)$.

Of course, (66) follows from (68). Now suppose the conversation turns to Ralph's beliefs about *the man he sees on the beach*, who unbeknownst to Ralph is Ortcutt. At this point, we might wonder whether in uttering (67) you had in mind the full strength of (65), from which it follows that Ralph does not believe of the man he sees on the beach—via the role $\mathbf{r}_b$, say—that he is a spy. The coherence of wondering this suggests that it is not a problematic result that (65) and (66) are not equivalent.

While Aloni's framework requires both the existence and uniqueness conditions for any fixed set F of functions, it also allows that different sets of functions may serve as the domain of quantification in different contexts. To formalize this idea in the style of Aloni (2005, §5) with a language of z variables, expand the set of terms to include for all $i \in \mathbb{N}$ a set of variables $\{z^i_1, z^i_2, \dots\}$ associated with context i; second, redefine a model to be a tuple $\mathscr{M} = \langle W, R_a, R_d, D, \mathbf{F}, V\rangle$ such that $\mathbf{F} \subseteq \mathscr{P}(D^W)$ and for all $F \in \mathbf{F}$, $\mathscr{M} = \langle W, R_a, R_d, D, F, V\rangle$ satisfies Definition 2; third, redefine an assignment to be a function π like μ in Definition 3 but extended so that for all $i \in \mathbb{N}$, $\pi(i) \in \mathbf{F}$; finally, redefine the clause for quantification with z variables:

$$\mathscr{M}, w \vDash_\pi \forall z^i_j \varphi \text{ iff for all } f \in \pi(i),\ \mathscr{M}, w \vDash_{\pi[z^i_j/f]} \varphi.$$

Hence for each context i, z^i variables are associated with their own domain of quantification $\pi(i) \in \mathbf{F}$.[30] We can recover the semantics of §3 by requiring that $|\mathbf{F}| = 1$.

As suggested at the end of §3, one can easily generalize the FOIL semantics to allow many sets of roles, only we would use superscripts on role variables.

[30] Aloni considers it an advantage of this more general semantics that we can have

(69) $\mathscr{M}, w \vDash_\pi \exists z^i \varphi(z^i) \wedge \neg\exists z^j \varphi(z^j)$,

as if there is a shift in context mid-formula. Instead of doing this with one of Aloni's models, we could consider two regular models $\mathscr{M} = \langle W, R_a, R_d, D, \pi(i), V\rangle$ and $\mathscr{M}' = \langle W, R_a, R_d, D, \pi(j), V\rangle$, each associated with a different context, such that

(70) $\mathscr{M}, w \vDash_\mu \exists z \varphi(z)$ and $\mathscr{M}', w \vDash_{\mu'} \neg\exists z \varphi(z)$.

Aloni's motivation for considering (69) is the following kind of reasoning:

(I) Ralph believes that the man with the brown hat is a spy.
(II) The man with the brown hat is Ortcutt.
(III) So Ralph believes of Ortcutt that he is a spy.

(IV) Ralph believes that the man seen on the beach is not a spy.
(V) The man seen on the beach is Ortcutt.
(VI) So Ralph does *not* believe of Ortcutt that he is a spy.

Aloni concludes that

(71) $\exists z^1 (z^1 = o \wedge \boxed{d} S(z^1)) \wedge \neg\exists z^2 (z^2 = o \wedge \boxed{d} S(z^2))$

should be satisfiable, which it is in her semantics. However, it seems to us to be a mistake to conclude (VI) on the basis of (IV) and (V). Instead, by analogy with (I) - (III), one should conclude

(VI$'$) So Ralph believes of Ortcutt that he is not a spy.

Then we can express the compatibility of (III) and (VI$'$) by the satisfiable sentence

(72) $\exists z_1 (z_1 = o \wedge \boxed{d} S(z_1)) \wedge \exists z_2 (z_2 = o \wedge \boxed{d} \neg S(z_2))$.

This is not to say, however, that there are not other good motivations for the more general semantics.

6 Multiple Agents and Points of View

In distinguishing roles from individual concepts, we emphasized the *agent-relativity* of roles. We will now make this relativity more explicit by extending our framework to a multi-agent language and multi-agent models. Typically the move from single to multi-agent epistemic logic is a matter of subscripting operators and relations by agent labels. We will take a different approach in two ways. First, instead of introducing many doxastic operators, we will introduce many "point of view" operators. Second, instead of subscripting these operators with agent labels, we will subscript them by *terms* of our language, so that agents will be individuals in our domain.

Definition 12 (Multi-Agent Language). Given the same sets of basic symbols as in Definition 1, the multi-agent language is generated by the grammar ($i,k \in \mathbb{N}$):

$$t \ ::= n_i \mid x_i$$
$$\varphi \ ::= Q_i^k(t_1,\ldots,t_k) \mid t = t' \mid \mathsf{P}(t,r_i) \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \boxed{a}\varphi \mid \boxed{d}\varphi \mid \forall x_i\varphi \mid \forall r_i\varphi \mid \mathsf{pov}_t\varphi.$$

The only addition are the "point of view" operators pov_t, with the intended reading:

$\mathsf{pov}_t\varphi$ "from the point of view of t, φ"
(or more technically, "centering on t, φ").

The unsubscripted P predicate and $\boxed{d}$ operator can now be read as:

$\mathsf{P}(t,r_i)$ "t plays role r_i (for the agent at the center)";
$\boxed{d}\varphi$ "it is doxastically necessary (for the agent at the center) that φ."

Thus, intuitively, $\mathsf{pov}_t\boxed{d}\varphi$ indicates that *for individual t, it is doxastically necessary that φ*; and $\mathsf{pov}_t\mathsf{P}(t',r_i)$ indicates that *for individual t, t' plays role r_i*.[31] Note that allowing *variables* to occur as subscripts of the point of view operators significantly increases the language's expressive power, since these variables may then be bound by quantifiers. As a simple example, in this framework (as in that of Grove 1995) one can express the likes of "someone believes φ," using $\exists x\, \mathsf{pov}_x\boxed{d}\varphi$.

Definition 13 (Multi-Agent Model). A *multi-agent* constant domain model for the language of Definition 12 is a tuple $\mathscr{M} = \langle W, R_a, \mathrm{R}_d, D, \mathrm{F}, V\rangle$ such that:

- W, R_a, D, and V are as in Definition 2;
- R_d is a binary relation on $W \times D$;
- F is a set of partial functions from $W \times D$ to D.

The only difference between these models and those of Definition 2 is that the doxastic relation R_d and role functions in F now apply to Lewis's (1979) "centered

[31] In English, to say "from the point of view of t, φ," might suggest that *t believes φ*, but this it not the intended reading of $\mathsf{pov}_t\varphi$, as its formal truth definition below makes clear.

worlds," which are pairs of a possible world and an individual—the center.[32] For the doxastic relation, given any $w, w' \in W$ and $c, c' \in D$, we take $\langle w, c\rangle R_d \langle w', c'\rangle$ to mean that *it is compatible with individual c's beliefs in w that she is individual c' in w'* (as in Stalnaker 2010, 70), allowing an agent to have uncertainty both about the world and about *her own identity*. Since the relevant agent is given by the center c of the first centered world $\langle w, c\rangle$, we only need one R_d relation to represent the beliefs of multiple agents. Finally, for the roles, given any $f \in \mathrm{F}$, $w \in W$, and $c \in D$, we take $f(w, c)$ to be the object that plays role f in world w *for the individual c*.

A variable assignment μ now maps role variables r_i to elements of F.

Definition 14 (Truth). Given a multi-agent model $\mathscr{M}$, $w \in W$, $c \in D$, assignment μ, and formula φ, define $\mathscr{M}, w, c \vDash_\mu \varphi$ as follows (with other clauses as in Def. 5):

$$
\begin{array}{lll}
\mathscr{M}, w, c \vDash_\mu \mathsf{P}(t, r_i) & \text{iff} & \mu(r_i) \text{ is defined at } \langle w, c\rangle \text{ and } [t]_{\mathscr{M}, w, \mu} = \mu(r_i)(w, c); \\
\mathscr{M}, w, c \vDash_\mu \boxed{a}\varphi & \text{iff} & \text{for all } w' \in W, \text{ if } wR_a w' \text{ then } \mathscr{M}, w', c \vDash_\mu \varphi; \\
\mathscr{M}, w, c \vDash_\mu \boxed{d}\varphi & \text{iff} & \text{for all } w' \in W, c' \in D, \text{ if } \langle w, c\rangle R_d \langle w', c'\rangle \\
 & & \qquad \text{then } \mathscr{M}, w', c' \vDash_\mu \varphi; \\
\mathscr{M}, w, c \vDash_\mu \mathsf{pov}_t\varphi & \text{iff} & \mathscr{M}, w, [t]_{\mathscr{M}, w, \mu} \vDash_\mu \varphi.
\end{array}
$$

Hence pov_t can be thought of as a "center shifting" operator. Like the similar $@_s$ operators of hybrid logic (see Areces and ten Cate 2006), the pov_t operators are normal modal operators validating the K axiom ($\mathsf{pov}_t(\varphi \to \psi) \to (\mathsf{pov}_t\varphi \to \mathsf{pov}_t\psi)$) and necessitation rule (if φ is valid, so is $\mathsf{pov}_t\varphi$), plus self-duality ($\mathsf{pov}_t\varphi \leftrightarrow \neg\mathsf{pov}_t\neg\varphi$). One can define the agent-indexed $\boxed{d}_t\varphi := \mathsf{pov}_t\boxed{d}\varphi$ and $\mathsf{P}_t(t', r_i) := \mathsf{pov}_t\mathsf{P}(t', r_i)$, but there are logical reasons not to take these defined operators as primitive.[33]

Let us begin by analyzing Perry's (1977) case of the man Heimson who believes he is Hume, with an added twist of "double vision." Suppose that Hume plays two roles in Heimson's life: first, Hume is the author of the books labeled 'David Hume' on Heimson's shelf, a role we will assign to the role variable r_a (for *a*uthor); second, Hume is a pen pal of Heimson's (here taken to be a contemporary of Hume) who signs his letters with the name 'D.H.', a role we will assign to r_{pp} (for *p*en *p*al). Finally, Heimson is the person who Heimson finds out about by introspection, proprioception, etc., playing the *self-role* that we will assign to r_{self}.[34] The catch is that Heimson is confused about his own identity in such a way that we can say:

(73) Heimson believes that he is Hume and that D.H. lives far away.

---

[32] While we take a centered world to be any element of $W \times D$, one may wish to only admit pairs $\langle w, c\rangle$ such that c is an agent (in some distinguished set $Agt \subseteq D$) and c exists in w (using the existence predicate of §3, $c \in V(\mathsf{E}, w)$), but for simplicity we do not make these assumptions here.

[33] By the truth definition, we have $\mathscr{M}, w, c \vDash_\mu \boxed{d}_t\varphi$ iff for all $w' \in W$, $c' \in D$, if $\langle w, [t]_{\mathscr{M}, w, \mu}\rangle R_d \langle w', c'\rangle$, then $\mathscr{M}, w', c' \vDash_\mu \varphi$. The problem with taking the $\boxed{d}_t$ operators as primitive instead of $\boxed{d}$ is that we would then lose important results of modal correspondence theory. For example, requiring that R_d be *reflexive* (thinking of it now as an *epistemic* accessibility relation) would not guarantee the validity of $\boxed{d}_t\varphi \to \varphi$, since the reflexivity of R_d would not guarantee that $\langle w, [t]_{\mathscr{M}, w, \mu}\rangle R_d \langle w, c\rangle$. But reflexivity would guarantee the validity of $\boxed{d}\varphi \to \varphi$, as desired.

[34] One may want to define the self-role such that for all worlds w and agents c, $f_{self}(w, c) = c$.

We can formalize (73) as follows:[35]

(77) $\mathsf{pov}_{Heimson}\big[\mathsf{P}(Heimson, r_{self}) \wedge \mathsf{P}(Hume, r_a) \wedge \mathsf{P}(D.H., r_{pp}) \wedge$
$\boxed{d}\exists x \exists y \exists z(\mathsf{P}(x, r_{self}) \wedge \mathsf{P}(y, r_a) \wedge \mathsf{P}(z, r_{pp}) \wedge x = y \wedge FA(z))\big].$

Let us now see how this framework allows us to analyze multi-agent belief ascriptions. Building on Example 1 from §4.3, suppose that the registrar is looking at Egbert and gathers from Egbert's grin that he thinks there is a fish on the hook:

(78) The registrar believes that Egbert believes there is a fish on the hook.

Where r_s is assigned the role of *being the person seen*, we can translate (78) as:

(79) $\mathsf{pov}_{registrar}\big[\mathsf{P}(Egbert, r_s) \wedge \boxed{d}\exists x_2(\mathsf{P}(x_2, r_s) \wedge \mathsf{pov}_{x_2}\boxed{d}\exists x_3 FOH(x_3))\big]$

We take it that (78) entails:

(80) The registrar believes that someone believes there is a fish on the hook.

With a de re reading, we formalize (80) as

(81) $\mathsf{pov}_{registrar}\exists x_1 \exists r\big[\mathsf{P}(x_1, r) \wedge \boxed{d}\exists x_2(\mathsf{P}(x_2, r) \wedge \mathsf{pov}_{x_2}\boxed{d}\exists x_3 FOH(x_3))\big].$

Many other interesting multi-agent belief ascriptions can be handled in this way. Let us look at one more famous example, due to Mark Richard (1983). A man *m* sees a woman *w* in a phone booth. As he watches her from his office window, he sees that an out-of-control steamroller is headed toward the phone booth. He waves wildly to warn her. At the same time, he is talking on the phone to a friend. She tells him of a strange man who is waving wildly to her, apparently believing she is in danger. Of course, unknown to *m*, he is talking to the very woman he his seeing, without realizing it. In this case, *m* might tell *w* over the phone, "I believe you are not in danger," while at the same time agreeing with her that "The man waving at you believes you are in danger." How is this coherent? The answer from Crimmins and

[35] A similar analysis applies to other well-known problems in the theory of reference, such as Castañeda's (1966) puzzle about the first person. Surely through most of his life after 1884, Samuel Clemens believed that he wrote *Huckelberry Finn*. But one can imagine that in his dotage, Clemens held a copy of the book in his hand, saw that it was written by Mark Twain, but couldn't remember that 'Mark Twain' had been his pseudonym and had no inclination to say "I wrote this." Castañeda made the point, with many similar examples, that even in the latter case, we *could* say

(74) Samuel Clemens believes that he wrote *Huckelberry Finn*.

since he is Mark Twain, and he believes that Mark Twain wrote *Huckelberry Finn*. However, in the sense in which it was true through much of his life that he believed he wrote *Huckelberry Finn*, at this moment late in his life, it is not. There is a reading of (74) on which it is false.

In the case we are imagining, Samuel Clemens plays (at least) two roles in Samuel Clemens' life, the self-role $\mathbf{r}_{self}$ and the role $\mathbf{r}_{MT}$ of being the source of the 'Mark Twain' name-network that is exploited by the use of that name on the book he holds in his hands. Given this, we can distinguish between the two readings of (74), the first false and the second true, as follows:

(75) $\mathsf{pov}_{Samuel}[\mathsf{P}(Samuel, r_{self}) \wedge \boxed{d}\exists x(\mathsf{P}(x, r_{self}) \wedge Wrote(x, HF))];$
(76) $\mathsf{pov}_{Samuel}[\mathsf{P}(Samuel, r_{MT}) \wedge \boxed{d}\exists x(\mathsf{P}(x, r_{MT}) \wedge Wrote(x, HF))].$

Perry (1989) is that the choice of words in the *subject* position of the belief reports ('I' or 'the man waving at you') can affect what is the relevant *role* via which the subject is said to believe something of the object: the first belief attribution is true iff m believes of w via the role $\mathbf{r}_{phoned}$ that she is not in danger, while the second is true iff m believes of w via the role $\mathbf{r}_{seen}$ that she is in danger. Mapping variables r_{phoned} and r_{seen} to these roles, we can describe m's doxastic state as follows:

(82) $\mathsf{pov}_m\big[\mathsf{P}(w, r_{phoned}) \wedge \boxed{d}\exists x_1(\mathsf{P}(x_1, r_{phoned}) \wedge \neg \mathit{InDanger}(x_1))\big]$;

(83) $\mathsf{pov}_m\big[\mathsf{P}(w, r_{seen}) \wedge \boxed{d}\exists x_1(\mathsf{P}(x_1, r_{seen}) \wedge \mathit{InDanger}(x_1))\big]$;

(84) $\mathsf{pov}_m\big[\mathsf{P}(w, r_{phoned}) \wedge \boxed{d}\exists x_1\big(\mathsf{P}(x_1, r_{phoned}) \wedge \neg \mathit{InDanger}(x_1) \wedge$
$\exists ! x_2\big(\mathit{WavingAt}(x_2, x_1) \wedge$
$\mathsf{pov}_{x_2}\big[\mathsf{P}(x_1, r_{seen}) \wedge \boxed{d}\exists x_3(\mathsf{P}(x_3, r_{seen}) \wedge \mathit{InDanger}(x_3))\big]\big)\big)\big]$.

We leave it to the reader to further explore the possibilities for formalizing multi-agent belief ascriptions in this framework. We also leave it to future work to investigate the differences between FOIL and the logic over our multi-agent models.

7 Conclusion

We began this paper with the idea that epistemic predicate logic faces a problem that alethic predicate logic does not: the problem of the cognitive fix. As we saw, this problem requires a different solution than the solution, based on modally loyal names, to the problems that Quine raised for alethic predicate logic. However, we have argued that the solution to the problem of the cognitive fix is not to treat names, worlds, or individuals differently in epistemic logic than in alethic logic. The solution does not require giving up what we called the Conservative Approach.

Instead, the solution requires giving up the idea that translating belief ascriptions into modal logic follows the simple pattern of translating necessity claims, what we called the Complement = Operand Hypothesis. We argued for an alternative approach to formalizing belief reports, based on making explicit the *unarticulated constituents* of such reports. Taking these unarticulated constituents to be the *roles* that the objects of belief play in the cognitive life of the believer, we carried out the formalizations in a version of Fitting's First-Order Intensional Logic. We applied the idea of agent-relative roles to the Hintikka-Kripke Problem for alethic-epistemic logic, to quantification into epistemic contexts, and to multi-agent belief ascriptions.

The move from individual concepts to agent-relative roles also opens up new ways of thinking about the dynamics of knowledge and belief. One can easily add to our framework the basic machinery of dynamic epistemic logic (van Benthem, 2011): when an agent learns φ, update the model by cutting doxastic/epistemic accessibility links to $\neg\varphi$-worlds. But now we can consider not only the dynamics of the relations, but also the *dynamics of roles*. When an agent makes an observation of the world, it is not just that she receives information that rules out epistemic possibilities; in addition, the objects that she observes come to play various roles in her cognitive life, making it possible for her to have new thoughts about those objects. A dynamic epistemic predicate logic that allows the accessibility relations R_d to change should not freeze the set F of functions in place. While Fregean senses may be static, agent-relative roles are not. Perhaps it is not only the dynamics of *ruling out*, but also the dynamics of *roles* that belongs on the agenda of logical dynamics.

Acknowledgements For helpful discussion or comments on this paper, we thank Johan van Benthem, Russell Buehler, Thomas Icard, David Israel, Ethan Jerzak, Alex Kocurek, Daniel Lassiter, John MacFarlane, Michael Rieppel, Shane Steinert-Threlkeld, Justin Vlasits, and Seth Yalcin.

References

Aloni M (2005) Individual Concepts in Modal Predicate Logic. Journal of Philosophical Logic 34(1):1–64

Areces C, ten Cate B (2006) Hybrid Logic. In: Blackburn P, van Benthem J, Wolter F (eds) Handbook of Modal Logic, Elsevier, pp 821–868

Arló-Costa H, Pacuit E (2006) First-Order Classical Modal Logic. Studia Logica 84:171–210

Barnes KT (1976) Proper Names, Possible Worlds, and Propositional Attitudes. Philosophia 6(1):29–38

Barwise J, Perry J (1983) Situations and Attitudes. MIT Press

Belardinelli F, Lomuscio A (2009) Quantified epistemic logics for reasoning about knowledge in multi-agent systems. Artificial Intelligence 173:982–1013

van Benthem J (1985) Modal Logic and Classical Logic. Bibliopolis, Napoli

van Benthem J (1993) Beyond Accessibility: Functional Models for Modal Logic. In: de Rijke M (ed) Diamonds and Defaults, Kluwer Academic Publishers, Dordrecht, pp 1–18

van Benthem J (2010a) Frame Correspondences in Modal Predicate Logic. In: Feferman S, Sieg W (eds) Proofs, Categories and Computations: Essays in Honor of Grigori Mints, College Publications, London

van Benthem J (2010b) Modal Logic for Open Minds. CSLI Publications, Stanford, CA

van Benthem J (2011) Logical Dynamics of Information and Interaction. Cambridge University Press

Braüner T, Ghilardi S (2006) First-Order Modal Logic. In: Blackburn P, van Benthem J, Wolter F (eds) Handbook of Modal Logic, Elsevier, Amsterdam, pp 549–620

Burge T (1973) Reference and Proper Names. The Journal of Philosophy 70(14):425–439

Carlson L (1988) Quantified Hintikka-Style Epistemic Logic. Synthese 74(2):223–262

Carnap R (1947) Meaning and Necessity. University of Chicago Press

Castañeda HN (1966) 'He': A Study in the Logic of Self-Consciousness. Ratio 7:130–157

Cresswell MJ, von Stechow A (1982) *De Re* Belief Generalized. Linguistics and Philosophy 5:503–535

Crimmins M (1992) Talk about Belief. MIT Press

Crimmins M, Perry J (1989) The Prince and the Phone Booth: Reporting Puzzling Beliefs. Journal of Philosophy 86(12):685–711

Cumming S (2008) Variabilism. Philosophical Review 117(4):525–554

Davies M, Humberstone L (1980) Two Notions of Necessity. Philosophical Studies 38:1–30

Etchemendy J (1999) The Concept of Logical Consequence. CSLI Publications, Stanford, CA

Fitting M (2004) First-order intensional logic. Annals of Pure and Applied Logic 127(1-3):171193

Fitting M (2006) FOIL Axiomatized. Studia Logica 84(1):1–22

Fitting M, Mendelsohn RL (1998) First-Order Modal Logic. Kluwer, Dordrecht

Føllesdal D (1961) Referential Opacity and Modal Logic. Routledge, published 2004.

Frege G (1892) Über Sinn und Bedeutung. Zeitschrift für Philosophie und philosophische Kritik 100:25–50
Garson JW (2001) Quantification in Modal Logic. In: Gabbay D, Guenthner E (eds) Handbook of Philosophical Logic, vol 3, 2nd edn, Kluwer, Dordrecht, pp 267–323
Garson JW (2006) Modal Logic for Philosophers. Cambridge Universtiy Press
Grove AJ (1995) Naming and identity in epistemic logic Part II: a first-order logic for naming. Artificial Intelligence 74:311–350
Hintikka J (1962) Knowledge and Belief: An Introduction to the Logic of the Two Notions. College Publications, London, republished 2005.
Hintikka J (1969) On the Logic of Perception. In: Models for Modalities, Reidel, Dordrecht, pp 151–183
Hintikka J (1970) Objects of Knowledge and Belief: Acquaintances and Public Figures. Journal of Philosophy 67:869–883
Holliday WH (2013a) Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism. Journal of Philosophical Logic forthcoming.
Holliday WH (2013b) Epistemic Logic and Epistemology. In: Hansson SO, Hendricks VF (eds) Handbook of Formal Philosophy, Springer, forthcoming.
Hughes GE, Cresswell MJ (1996) A New Introduction to Modal Logic. Routledge
Israel D, Perry J (1996) Where Monsters Dwell. In: Seligman J, Westerståhl D (eds) Logic, Language and Computation: The 1994 Moraga Proceedings, CSLI Publications, pp 303–16
Kaplan D (1968) Quantifying In. Synthese 19:178–214
Kaplan D (1989) Demonstratives. In: Almog J, Perry J, Wettstein H (eds) Themes from Kaplan, MIT Press, pp 481–563
Kripke S (1980) Naming and Necessity. Cambridge, Harvard University Press
Lehmann S (1978) The Hintikka-Kripke Problem. Philosophia 8:59–70
Lewis D (1977) Possible-World Semantics for Counterfactual Logics: A Rejoinder. Journal of Philosophical Logic 6:359–363
Lewis D (1979) Attitudes De Dicto and De Se. Philosophical Review 88(4):513–543
Lewis D (1983) Individuation by Acquaintance and by Stipulation. Philosophical Review 92(1):3–32
Linsky L (1979) Believing and Necessity. Theory and Decision 11:81–94
Ninan D (2012) Counterfactual attitudes and multi-centered worlds. Semantics and Pragmatics 5(5):1–57
Parks Z (1974) Semantics for Contingent Identity Systems. Notre Dame Journal of Formal Logic 15(2):333–334
Perry J (1977) Frege on Demonstratives. Philosophical Review 86:474–497
Perry J (1986) Thought without Representation. Proceedings of the Aristotelian Society 60:137–151
Perry J (2009) Hintikka on Demonstratives. Revue internationale de philosophie 4(250):369–382
Perry J (2012) Reference and Reflexivity, 2nd edn. CSLI Publications, Stanford, CA
Perry J (2013) Føllesdal and Quine's Slingshot. In: Frauchiger M (ed) Reference, Rationality, and Phenomenology: Themes from Føllesdal, Ontos Verlag, pp 237–258
Priest G (2002) The Hooded Man. Journal of Philosophical Logic 31(5):445467
Quine W (1953) Three Grades of Modal Involvement. In: Proceedings of the 11th International Congress of Philosophy, North-Holland Publishing, vol 14, pp 158–176
Quine W (1956) Quantifiers and Propositional Attitudes. The Journal of Philosophy 53(5):177–187
Rabinowicz W, Segerberg K (1994) Actual Truth, Possible Knowledge. Topoi 13:101–105
Richard M (1983) Direct Reference and Ascriptions of Belief. Journal of Philosophical Logic 12:425–452
Russell B (1905) On Denoting. Mind 14(4):479–493
Russell B (1911) Knowledge by Acquaintance and Knowledge by Description. Proceedings of the Aristotelian Society 11:108–128
Schroeter L (2012) Two-dimensional semantics. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy, winter 2012 edn

Soames S (1989) Direct Reference and Propositional Attitudes. In: Almog J, Perry J, Wettstein H (eds) Themes from Kaplan, MIT Press, pp 393–419
Stalnaker RC (1984) Inquiry. MIT Press
Stalnaker RC (2010) Our Knowledge of the Internal World. Oxford University Press
Wettstein H (1988) Cognitive Significance without Cognitive Content. Mind 97(385):1–28
Wishon D (2012) Russellian Acquaintance without Discriminating Knowledge, manuscript.
Yalcin S (2012) De Re from a Fregean Perspective, manuscript.