COGNITIVE SCIENCE **10**, 319–364 (1986)

Imputations and Explications: Representational Problems in Treatments of Propositional Attitudes

JOHN A. BARNDEN

Computer Science Department
Indiana University
Bloomington, Indiana 47405

The representation of propositional attitudes (beliefs, desires, etc.) and the analysis of natural-language, propositional-attitude reports presents difficult problems for cognitive science and artificial intelligence. In particular, various representational approaches to attitudes involve the incorrect "imputation," to cognitive agents, of the use of artificial theory-laden notions. Interesting cases of this problem are shown to occur in several approaches to attitudes. The imputation problem is shown to arise from the way that representational approaches explicate properties and relationships, and in particular from the way they explicate propositional attitudes themselves. Another factor contributing to imputation is the compositional nature of typical semantic approaches to propositional-attitude reports. Some strategies for avoiding undesirable imputation are examined. One of the main conclusions is that the importance of imputations that arise in a representation scheme depends strongly on the use to which the scheme is put—on whether, for instance, the scheme is used as part of a formal, objective account of natural language, or is used rather as a representational tool within an agent.

1. INTRODUCTION

Natural-language sentences about propositional attitudes, including the beliefs, hopes, desires, wonderings, doubts, etc. of cognitive agents, are generally agreed to present some of the most difficult problems in natural language semantics (see e.g., Barwise & Perry, 1983, and the introduction to Peters & Saarinen, 1982). Natural language aside, propositional attitudes are recognized as being of great importance in the theory of representation in philosophy and artificial intelligence (see e.g., Creary, 1979; Fagin & Halpern, 1985; Fodor, 1978; Hintikka, 1983; Konolige, 1984; Konolige,

I am grateful for advice and criticism from H.-N. Castañeda, R. Clark, J.M. Dunn, S.W. Kitchel, M.W. Lugowski, E.J. Smythe, and R.B.H. Stark, all at Indiana University, and from anonymous referees.

Correspondence and requests for reprints should be sent to the author at the address listed above.

1985; Levesque, 1984; Linsky, 1983; Maida & Shapiro, 1982; Rapaport & Shapiro, 1984). Propositional attitudes are the subjects of an immense body of literature in philosophy, theoretical linguistics, artificial intelligence, and cognitive science, and present many theoretical problems. Let us go straight to a simple illustration of the particular issue this paper is concerned with.

Consider some arbitrary body of water. Suppose we agree that the water is boiling if and only if it is forcibly expelling water vapor. In spite of this, the following sentences mean different things to us, and we can imagine assigning different truth values to them:

(S) *Mike believes that the water is boiling.*

(S′) *Mike believes that the water is forcibly expelling some water vapor.*

At least, the stated independence holds if the sentences are given certain natural interpretations. These interpretations take the complement—the clause following *believes that*—to portray faithfully the structure of Mike's thought in each case. Thus, (S′) takes Mike's belief to be based on concepts of expulsion, water vapor, and so on. On the other hand, (S) takes Mike's belief to be based on a concept of boiling. Then, provided we allow the possibility that Mike's concept of boiling is *not* such as to give the same structure to both beliefs, we have the freedom to say that one of (S), (S′) is true and the other false.

All this is familiar: It falls under the general heading of *opacity* of belief-complements in sentences. The type of opacity most discussed is *referential* opacity, whereby the replacement of a noun phrase in a belief-report complement by a co-referring phrase can change the truth value of the sentence. Thus, the sentences "John believes that Mary is clever" and "John believes that Jim's wife is clever" can have different truth values, even in the presence of the background fact that Mary is Jim's wife. But referential opacity is just a special case of a more wide-ranging opacity phenomenon, whereby the truth value of a belief-report can be changed by replacing the complement by a complement that follows from it (with the aid of background facts). It is this more general phenomenon that is at work in the (S), (S′) example above.

Consider now a hypothetical AI program, X, that is meant to be able to "understand" English. Program X, when given a sentence as input, renders it as a structure in some internal representation scheme RS that X uses to represent the world. RS is a predicate logic containing a modal belief operator B. (For a discussion of such operators, see e.g., Halpern and Moses [1985].) Suppose that (in a particular context) X renders *the water is boiling* as

(F) $(\exists v)\ (\text{forcibly-expel}(w,v) \wedge \text{is-water-vapor}(v))$.

Thus, X applies an *explicative*-rendering process with respect to *is boiling*. This approach by X is, let us assume, adequate for sentences about water boiling that do not talk about beliefs or other propositional attitudes.

But what would X's rendering of (S) be? A typical suggestion would be to adopt a *compositional* approach, whereby the rendering would be built up from the rendering of the complement. So, assume that X renders (S) as

(G) **B(mike, (∃v) (forcibly-expel(w,v) ∧ is-water-vapor(v))).**

The trouble with this is that it is also, surely, the natural rendering for X to construct for (S′), assuming that this sentence is interpreted in the structurally faithful way adopted earlier. Therefore, X's supposedly natural approach to the rendering of the two sentences leads to the same internal formula. This is unfortunate, as we have already seen that we can give different truth values to the sentences.

At the very least we can say that X confuses two sentences. But we can go further. Formula (G) is actually *more* natural as a rendering of (S′) than of (S): for the obvious reason that it conforms in structure more closely to the latter sentence. It is therefore reasonable to say that X, in rendering (S) as (G), is guilty of *erroneously imputing* to Mike a belief cast in terms of the concepts of expulsion, water vapor, and so on.

This example requires further discussion, which we defer until Section 3. The moral is that an explicative feature of a rendering scheme can seem adequate as long as belief reports are not considered, but becomes dubious when such reports are brought into the picture. The dubiousness consists in introducing a risk of imputing possibly erroneous belief-structures to agents. It is this issue of imputation that is the topic of this paper.

Much of our attention will be focused on imputations arising from explications of propositional attitudes themselves, rather than on explications of boiling and other earthy matters. So, we shall be considering *nested* belief reports of the form *Z believes that Y believes that*. . . .The imputations arising from proposed explications of the inner *believes that*. . . will be studied.

Parts of the paper, including the discussion of Barwise and Perry's system, involve a shift of viewpoint—away from considering the machinations of AI programs (and human beings) and towards considering "objective" accounts of propositional attitudes and propositional-attitude reports. By objective accounts, we mean those that are not focused on how a particular agent (program or human being) represents the propositional attitudes of agents. One general conclusion of the paper will be that this viewpoint distinction must be carefully attended to in the propositional-attitude field.

We shall frequently use the strategy applied to sentence (S). That is, we shall consider a belief report (S$_1$) and a formula (G$_1$) that is a putative rendering of it in some formal representation scheme; we shall then show that (G_1) is more naturally to be regarded as a rendering of a report (S'_1) that is independent in truth value from (S_1). Sentence (S'_1) will commonly be (S_1) with some part of its complement replaced by another phrase. Hence, such an argument will show that the rendering of (S_1) as (G_1) involves in effect an invalid substitution in the complement. Because of this relationship to the issue of opacity, our discussion is connected to much contemporary work on propositional attitudes. However, our deliberations appear to go beyond the studies that have been made before.

The remainder of the paper has the following structure:

- Section 2 explains some distinctions between different ways of using representation schemes.
- Section 3 looks at the boiling-water example more closely, and draws general conclusions about how "explication" can lead to possibly harmful inputations.
- Section 4 looks at an imputation problem that arises for nested-attitude reports in the "situation semantics" approach as reported in Barwise and Perry (1983). The section goes on to demonstrate imputation problems in the representation of nested propositional attitudes in the neo-Fregean approach of Creary (1979) and in quotational schemes.
- Section 5 considers two broad strategies for the reduction of imputation problems when designing representation schemes. The discussion leads into a sketch of a representation scheme I am developing. The scheme is strongly related to others that have been proposed: the point of mentioning it is to discuss the type of imputation it leads to.
- Section 6 discusses the role that the compositionality of approaches to attitude reports plays in generating imputations, and suggests a third strategy for imputation reduction: namely, eliminating compositionality.
- Section 7 briefly mentions the relationships of the imputation problem to some other issues that have been studied in the literature, for example, the translation of attitude reports between natural languages.
- Section 8 is the conclusion, and in particular makes the point that nothing in this paper is fundamentally tied to the use of *formal logic* as a representational medium.

2. DIMENSIONS OF PURPOSE

One of our themes is that, in the propositional-attitude field, one must take care to be clear about the different uses one might have in mind for a representation scheme. The categories of usage I concentrate on are as follows:

The OB/AG Distinction

A theoretician might put forward a particular representation scheme as a good way of describing the world or some aspect of it for his or her *own* theoretical purposes, without implying that any human or artificial cognitive agent does or should use the scheme in its cognitive processing. In this sense, the scheme is put forward as an "objective" (OB) way of representing the world.

On the other hand, in proposing a representation scheme one may be claiming that human agents actually use the scheme in their everyday cognition, or one may be proposing that an artificial cognitive agent (AI program) should use the scheme to represent its world. These are "AG" uses of the scheme. The scheme is, of course, supposed to be implemented somehow in the "hardware" of the brain or computer, but that is not our concern.

Note that if one has the AG purpose, then the scheme does not necessarily have to achieve the level of adequacy that one might demand for an OB scheme. A psychological theory might legitimately claim that human agents use a representation scheme that is deficient in some way. Equally, a representation scheme proposed for use in a particular AI program might not serve as a good OB scheme but might nevertheless be heuristically adequate in the program's environment. We shall be exploiting this difference between OB and AG uses.

The NL/GEN Distinction

Suppose a cognitive agent renders incoming natural language sentences as structures in an internal representation scheme. Then the scheme is being given an "NL" use (which is also an AG use). Equally, a philosopher or theoretical linguist might use a particular representation scheme as the medium in which to couch formal renderings of natural language sentences. By "renderings" I mean to include truth conditions, logical forms, semantic deep structures, "meanings," etc. The representation scheme is being given an NL use (that is also an OB use). In this case we also say that the theoretician is using the representation scheme as a basis for an *objective semantics.*

The formal rendering of a sentence in either the AG case or the OB case should take account of issues as the context of utterances of the sentence, but for brevity we shall be ignoring these pragmatic considerations.

If an AI program or human agent uses a scheme for general representational purposes (possibly, but not necessarily, including the rendering of inputed sentences), then the scheme has a GEN (and AG) use. Similarly, a theoretician might propose a particular scheme for his or her own theoretical purposes, with or without intending to use the scheme to render natural language sentences. The scheme then has a GEN (and OB) use.

The OB/AG distinction and the NL/GEN distinction are independent of each other. A use of a scheme can lie in any of the four quadrants induced (AG/NL, AG/GEN, OB/NL, OB/GEN), and a given scheme might well be subject to all four categories of use on different occasions.

3. IMPUTATION BY EXPLICATION: BASE CASE

The NL/AG Case

Consider again the example sentences from Section 1.

(S) *Mike believes that the water is boiling*

and

(S′) *Mike believes that the water is forcibly expelling some water vapor.*

In our discussion, we assumed that artificial agent X renders these on the basis of an explication of boiling in terms of forcible expulsion of water vapor. It might, however, be objected that the example is artificial, in that an AI program X would be likely to use instead an explication that is more in tune with what ordinary human speakers mean by *boiling*. The putative "objection" actually plays into our hands, since one of our points is precisely that problems of undesirable imputation can arise because explications of terms do not accord with ordinary speakers' meanings for them. Although our example may in itself be unrealistic, it serves to capture the essence of the imputation problem to be revealed in more interesting examples later.

It is instructive to see what happens if we suppose that X uses a more natural explication. Assume, by way of illustration, that X renders *The water is boiling* as the following formula:

(XF) $\textbf{bubbling(w)} \land \textbf{very-hot(w)} \land (\exists\textbf{m})\ (\textbf{white-mist(m)} \land \textbf{give-off(w,m)})$.

Accordingly, assume that X renders (S) as

(XG) $\textbf{B(mike, bubbling(w)} \land \textbf{very-hot(w)} \land$
$(\exists\textbf{m})\ (\textbf{white-mist(m)} \land \textbf{give-off(w,m)}))$.

Much as before, (XG) would be the natural one for X to use as its rendering of

(S″) *Mike believes that the water is bubbling, very hot, and giving off a white mist,*

assuming again that this sentence is interpreted in the structurally faithful way adopted in Section 1. Now, given our assumption that the explication of boiling that is being used in a natural one, it may well be that it is quite right for X to render (S) and (S″) as the same formula. That is, in rendering (S) as (XG), X does impute to Mike a belief couched in terms of bubbling and so on, but this imputation may be *correct* (benign). I say it *may be* correct, because there is still an opportunity for the imputation to be incorrect. It may, for instance, be that Mike has a conscious notion of boiling that does not explicitly bring in the notion of giving off a white mist (although it does perhaps involve notions of bubbling and of being very hot).

Another possibility is that on a particular occasion it may that Mike's entertaining a conscious notion of boiling may involve simply the use of internal speech featuring the word "boiling." Although Mike *could* go on, we might assume, to deduce that the water is bubbling and so on, it is possible also that he does not.

We may accept that X can *usually* get away with the imputation to Mike of a belief couched in terms of bubbling, etc. It is as if X performs a *plausible inference* about Mike, based on a mythical rule that says "when Mike thinks about water boiling he is thinking about it bubbling,. . .". However, a major lesson of modern AI research is that although the ability to make plausible inferences in very useful, even essential, there is also a need for a means to withdraw plausible inferences when the occasion arises. But X has no way of encoding a situation in which Mike believes that some water is boiling but is *not* entertaining notions of giving off a white mist and so on, assuming that X itself does not have an "is-boiling" predicate symbol and is stuck with using formulae like (XG) to express beliefs about boiling.

What sort of practical consequences could this inability on X's part have? Suppose X possesses the following inference rule:

(XR) When a person Z believes that there is a mist around and Z is in a room containing a painting, Z will remove the painting.

Assume that it is correct for X to deduce from (XG) that Mike believes that there is a mist around, and that X does so deduce on a particular occasion after being told sentence (S). So, if X knows Mike to be in a room containing a valuable painting, X will deduce that Mike will remove the painting. The trouble is, of course, that Mike may not actually believe that there is a mist around just because he believes that the water is boiling, since he may be so preoccupied that he has not made the necessary inference.

The moral is that a given type of explication can lead to in-principle erroneous imputations that *can* have important practical drawbacks, even though these particular imputations *usually* do no harm in practice. Observe that the imputations arise from the very act of explication, within a propositional-attitude context, rather than from any inherent incorrectness in the explication as such. It may be perfectly good for X to render the simple sentence *The water is boiling* in explicated form.

Nothing we have said relies crucially on X being artificial. Agent X could equally well have been human, although then we should no doubt take our assumption that X used a logic-based representation scheme as being merely for illustrative convenience.

The GEN/AG Case

Now we imagine a human or artificial cognitive agent X that is immersed in and represents a world containing other agents, but we turn attention away from X's processing of natural language input. The considerations presented just now under the NL/AG case can be carried over essentially intact. Instead of taking X to come to have information about Mike by virtue of processing sentences, we suppose that X has such information by virtue of its own inferencing and observations.

Assume that X's representation scheme contains nothing like a simple predicate symbol denoting the property of boiling, and that the scheme is forced to employ a formula like (XG) to represent situations in which Mike is thinking about boiling. Then, just the same sorts of imputation can arise as did before, and these imputations can be harmful in practice. For instance, the rule (XR) can be activated inappropriately. In fact, we are using the notion of erroneous imputation here as a convenient, intuitive way to make the point that the use of the explicative representation (XG) can lead X into making inappropriate inferences or performing inappropriate actions. Thus, our observations so far do not rest on natural language issues, but are perhaps easier to present in a natural language processing context.

If X's representation scheme does contain something like a predicate symbol for the property of boiling, then of course X can avoid making the imputations we have been examining, since X can entertain a formula like **B(mike, is-boiling(w))**.

The OB Case

The seriousness of the imputation problem increases when we move from the consideration of cognitive agents' ways of representing things, over to an examination of representation schemes used objectively.

Consider an attempt to provide an objective semantics for English explicating boiling as bubbling, being very hot and giving off a white mist. For example, consider a formal semantics that renders the sentence *The water is boiling* as formula (XG) in (modal) predicate logic, and assume that the logic contains no predicate symbol for boiling. Then, much as before, we may reasonably say the semantics is in danger of imputing an erroneous belief structure to Mike, since (XG) is more natural as a rendering of (S″), under the structurally faithful interpretation, than of (S).

Of course, this statement only makes sense if we allow for the possibility that there *is* a natural interpretation of (S) that is different from the assumed interpretation of (S″). This possibility arises if, for instance, it may be that Mike can deploy a simple representational item that stands for boiling. But then the trouble our hypothesized semantics is in is a result of its using a representation scheme that is impoverished with respect to the world. The particular difficulty would be avoided if this scheme itself contained a simple predicate symbol denoting boiling.

It may therefore seem that the imputation problem being raised is artificial, because all that the semanticist need do is to ensure that the logic used contains the boiling predicate symbol. This is of course true, but the examples we shall look at later involve forms of explication that are not so easy to avoid.

That said, there are influential semantic proposals in existence that do attempt to explicate natural language terms by means of representational primitives. For instance, in the semantic proposals of Schank (1973), all actions are meant to be *explicated* at the internal-representation level to complex representational structures built from just a few action primitives (and other sorts of primitive). It should be clear that, although our examples have involved the explication of a property rather than of an action directed from one object to another, similar imputation considerations will arise in the case of such actions.

Schank himself does not put his scheme forward as an objective semantics. Rather, he is interested in agent-based semantics: in how artificial and human agents process natural language. Nevertheless, our imputation considerations should act as a warning signal to theoreticians aiming at a Schankian objective semantics. And, the more primitive the proposed representational primitives become, the more likely is it that explications in terms of them will lead to *malign* imputations.

Going back to our consideration of an objective semantics that does render (S) as formula (XG), we have an imputation problem that is graver than that arising when an agent X so renders the sentence. This is because we no longer have appeal to the "heuristic argument" deployed in the NL/AG case. This argument was that the agent will usually be right in performing the imputation, even though on occasion it may be wrong. But our objective semantics is strictly incorrect if it imputes to Mike a belief couched in terms of bubbling, etc., *even if Mike usually* infers bubbling and so on from boiling.

To put this point differently, imputations that when performed by a *cognitive agent's* representation scheme are only occasionally harmful, are, if you like, *unqualifiedly* harmful when performed by an objective semantics using the same representation scheme.

Finally, the move we performed earlier from the NL/AG case to the GEN/AG case can be paralleled in the present OB context. A GEN/OB representation scheme is simply inadequate if it is the case that Mike can think about boiling without thinking about bubbling, etc., but the scheme's only tool for talking about Mike thinking about boiling is a formula like (XG).

A Cautionary Remark

One point needs to made clear about the sentences

(S) *Mike believes that the water is boiling*

(S″) *Mike believes that the water is bubbling, very hot, and giving off a white mist*

under structurally faithful interpretations. We have taken (S″) to convey that Mike is entertaining notions of bubbling, being-very-hot, giving-off, etc. Given this, there is a whole range of possible interpretations, based on different assumptions about the nature of these notions. For definiteness, let us fix on the assumption that the notions are Mike's *standard* notions of bubbling, etc. Still, this says nothing about the intrinsic nature of the notions themselves. In particular, the standardness assumption allows the notions to be arbitrarily complex.

But similar reasoning applies to (S). Under a structually faithful interpretation, which is what we use throughout, we take this sentence to convey that Mike is entertaining some notion of the water in question and his (standard) notion of the property of boiling. The latter notion can be arbitrarily complicated, and in particular it could conceivably be a complex built up from notions of bubbling, being-very-hot, giving-off, and so on. Thus, it would be a mistake to think that (S) definitely says that Mike is entertaining different notions from the ones he is entertaining according to (S″). The real difference between the two sentences is that (S″) definitely says that Mike is entertaining notions of bubbling, etc. (although the form of those notions is left unconstrained), whereas (S) merely allows the possibility that Mike is entertaining such notions. It is this difference of definiteness that is at the heart of the imputation issue.

If an agent X renders (S) as (XG), then X is imputing to Mike the entertainment of some (standard) notions or other of bubbling and so on, even though X may have no assumptions at all about the nature of those notions. If on the other hand X were to render (S) as **B(mike, boiling(w))**, then X is claiming merely that Mike is entertaining some (standard) notion of boiling. The nature of this notion is left unconstrained by the rendering; therefore, X *allows the possibility* that, but does not *claim* that, Mike is entertaining notions of bubbling, etc.

A Note on the Role of Consciousness

The concluding section briefly explains that the imputation issue is affected by the extent to which the having of a given propositional attitude is held to be a conscious phenomenon. I postpone the proper consideration of this matter, however, because of the great difficulties inherent in the notion of consciousness. Instead, I simply stipulate that all the beliefs mentioned in our discussion, or in sentences discussed, are *conscious*. (The believer will always be human.) In particular, in interpreting a propositional attitude report in a structurally faithful way, I assume that the agent has conscious notions of the objects, properties, and relationships mentioned in the complement. Thus, the structurally faithful interpretation of sentence (S″) takes Mike to have conscious notions of bubbling, being-white, mist, being-very-hot, and giving-off. Probably, the consciousness restriction is unnecessarily strong, and I intend to relax it in later studies, but it makes discussion of the issues more intuitively accessible.

When a human agent is consciously holding a belief, it may typically be the case that he or she is indulging in internal speech that expresses the content of the belief. Whether this is the case, however, is not our concern. Thus, in our discussion of the independence of sentences (S) and (S″), we do not *rely* on the idea that Mike *says to himself in English* the complements "the water is boiling" and "the water is bubbling, very hot, and giving off a white mist," although it may very well be the case that he does. Mike does not need to know English, or indeed to indulge in internal speech at all. Rather, we base the argument on the idea that Mike can have a conscious notion of boiling that does not explicitly bring in notions of bubbling, etc., without going into the question of what "having a conscious notion" means.

Section Summary

- Within propositional-attitude contexts, explicatory features of a representation scheme give rise to possibly harmful imputations.
- This is the case whether the representations scheme is being given an AG use or an OB use, and whether it is being given an NL use or a GEN use.
- In OB cases, imputations are more likely to be harmful than they are in AG cases, because a cognitive agent may usually be able to get away with making the imputations.
- An imputation is not necessarily malign. If in the above examples Mike *does* happen to think about boiling by means of notions of bubbling and so on, then the imputations are correct.
- Since imputation issues are brought out by considering the representation of propositional attitudes and the rendering of propositional attitude reports, a representation scheme that may seem adequate when propositional attitudes are not making an appearance may cease to look adequate when they do appear.

It is worth also making the methodological point that even when imputations are correct, the theorist who puts forward the particular representation scheme in question may not *realize* that the imputations can result. Therefore, the value of our analysis, even in the case when imputations are correct, is to increase the level of theoretical awareness of the imputation issue in general.

Although our examples have used a modal belief operator, the imputation problem is not induced by modal logic per se. Similar imputation phenomena arise when other types of logic are used. For example, nothing in our discussion changes if we replace modal formulae like **B(agent, P)** by quotational formulae like **B(agent, 'P').**

Imputations have been presented as being caused just by explications. As we shall see later, however, there are other contributory factors, and in some schemes there is a way of "packaging" explications so that they do not lead to imputations.

4. IMPUTATION BY EXPLICATION OF PROPOSITIONAL ATTITUDES

There is an especially interesting way in which explication can lead to harmful imputation. This is, loosely speaking, when what is explicated within a propositional-attitude context is itself a propositional attitude. We shall see that the harmful imputation arises in this way in "situation semantics" as reported in Barwise and Perry (1983)[1], in the neo-Fregean approach of Creary (1979), and in expression-denoting schemes (e.g., logics that can quote their own expressions).

Barwise and Perry's OB/NL Scheme

In this subsection, we appeal to grossly simplified forms of the ideas in Barwise and Perry's (1983) study. In our simplified version of their theory, sentences are taken to describe *situation*—roughly, partial states of affairs. Exactly what situations are is left largely to intuition in Barwise and Perry (1983), but, as a simple example, the sentence "Mary is clever" is a natural-language description of the situation of Mary's being clever. The sentence is true if the situation is "actual" in some sense.

Barwise and Perry provide a logic-like representation scheme for describing situations. Thus, if we use this scheme to describe the situations described by English sentences, we are "rendering" the sentences by means of the scheme and are thereby giving it an OB/NL use.

The sentence

(S1) *Mary is clever and beautiful*

describes the situation that is formally described by[2]

(BP1) **[clever(Mary); beautiful(Mary)].**

The sentence

(S2) *Mike is taller than Bill*

[1] It should be said that Barwise and Perry are in the process of modifying their treatment of propositional attitude reports (see Barwise and Perry [1985]). It remains to be seen whether the modified approach will avoid the imputation problems.

[2] The formal notation used in this section, as well as the general terminology, is a simplification of Barwise and Perry's.

describes the situation (formally described by)

(BP2) **[taller-than(Mike, Bill)].**

Our concern is with propositional attitude reports. For representing belief situations, Barwise and Perry introduce a relation B_r on agents and situations, to over-simplify a little. Roughly, believing agents are cast as being in B_r relationships to situations (the believed states of affairs). One might expect from this that the sentence

(S3) *Mike believes that Mary is clever and beautiful*

would be rendered as **[B_r(Mike, [clever(Mary); beautiful(Mary)])].** However, Barwise and Perry choose to use the more complex formula

(BP3) **[B_r(Mike, [clever(x); beautiful(x)]);**
of(x,Mary)].

for reasons that need not detain us. The expression

[clever(x); beautiful(x)],

describes a situation-type T rather than a situation. The situation-type is a parametrization of the situation described by (BP1). The belief-situation described by (BP3) is one in which Mike's "frame of mind" is the situation-type T. The intuition is that Mike has a certain *type* of belief, and that type is T. B_r thus relates cognitive agents to situation types rather than to situations. The "**of**" clause in (BP3) serves to bind (or "anchor") the parameter x to the person Mary.

In sum, (BP3) describes, in a somewhat roundabout way, the belief-situation described by sentence (S3). A belief-situation is a situation in which a cognitive agent has a frame of mind consisting of a situation type whose parameters are anchored to individuals. Barwise and Perry also talk in terms of the cognitive agents' mental state being *classified* by the situation-types.

To keep things simple, we shall gloss over the point that sentence (S3) is susceptible to multiple interpretations. For instance, there are interpretations that explicate *clever,* much as we explicated *boiling* earlier. In any case, Barwise and Perry do not discuss the explication of properties like being-clever—the only explications they deal with are those of propositional attitudes (and certain relationships that work in a similar way). So, we shall take (BP3) to be the "standard" rendering of (S3).

Consider now:

(S4) *George believes that Mike believes that Mary is clever.*

Although Barwise and Perry do not present a B_r-based account of nested belief reports, it is natural to extrapolate from their examples by taking (S4) to describe the situation

(BP4) **[B_r(George, [B_r(x, [clever(y)]);**
of (y,y′)]
of(x,Mike); of (y′,Mary)].

The trouble with postulating this situation as being described by (S4) is that we can regard it *more* naturally as being described by the following sentence:

(S4′) George believes that (i) Mike bears-the-$B_r$relationship to
the y-being-clever situation-type,
where (ii) y is anchored to Mary,

under a structurally faithful interpretation (one taking the complement to correspond faithfully to the structure of the belief). To treat the complement thus is quite in line with Barwise and Perry's examples.

We have of course augmented English with some new technical vocabulary, but presumably Barwise and Perry's theory is meant to be robust under vocabulary augmentations of whatever sort that might become convenient in discourse. In any case, we do not have to assume an augmented vocabulary. Instead of (S4′) we could have used:

(S4″) *George believes that Mike is in the relationship Barwise and Perry denote by 'B_r' to the parametrization of the entity they take to be described by the sentence "Mary is clever," where the parameter is linked to Mary by the relationship they denote by 'of'.*

This sentence is less convenient for the purposes of discussion, because not only is it more complex, but also requires a different sort of interpretation. The complement must not be taken in a structurally faithful way in its entirety; for instance, phrases like "the relationship they denote by 'B_r'" must be regarded merely as being for the benefit of the speaker and hearer (there being no commitment to the way that George thinks of the relationship).

It is highly undesirable that an objective semantic theory such as Barwise and Perry's should render (S4) and (S4′) in the same way. They would clearly mean different things to most people who understand the special vocabulary, and would probably be taken by most people to have different truth values; therefore, the theory should assign different renderings to them. It is conceivable that there are particular people (situation semanticists?) who would interpret both sentences to mean the same thing: but that is a side issue.

Notice that we have adopted the same strategy as was used in the analysis of sentence (S) in Section 1 and Section 3. What has happened is that Barwise and Perry's scheme gives *believes that Mary is clever* in (S4) an explication in terms of situation-types, anchoring and the belief-state classification (B_r) relation, in such a way that we can intuitively say that the scheme causes an imputation of these theoretical notions to George.

In saying that it is unreasonable to impute the notion of the B_r relationship to ordinary human agents, we must remember that we are talking about *conscious* notions. This is because we assume throughout this paper that the beliefs under discussion are all conscious. (Recall the Note near the end of Section 3). But we should also remember that we are not relying on the idea that an ordinary agent is unlikely to indulge in internal speech involving the terms "B_r relationship," "belief classification," or whatever. Rather, we are saying that the B_r relationship is simply something that an ordinary agent is not consciously aware of, whatever the form his/her conscious thoughts may take.

The sort of confusion illustrated by that of (S4) and (S4′) can be of practical importance, at a sophisticated level. We might want to represent a state of affairs in which (S4) is the case whereas (S4′) is not, and George is in a community of researchers studying Barwise and Perry's theory. Consider the following dialogue, for instance.

Jill: I was talking to George yesterday. He believes that Mike believes that Mary is clever.

Jack: Oh. . .so he believes that Mike bears-the-B_r-relationship to the y-being-clever situation-type where y is anchored to Mary.

Jill: No, George has never met the idea of your precious B_r relationship!

Jack: Hmm. But surely if George were a Barwise and Perry disciple, my statement would have been right?

Jill: Possibly. It's still possible though that he might not have actually gone to the trouble of deducing what you said he believes from his belief that Mike believes that Mary is clever.

A proper account of the meaning of this dialogue should surely allow Jill to say "No" in her second statement without contradicting herself.

It is important to realize that there is nothing in our considerations that contradicts the claim that a person believes something if and only if that person is in a B_r relationship to the appropriate situation-type, anchored in the appropriate way. What is being contradicted is the idea that this explication can, so to speak, be used within propositional-attitude contexts. The point is very similar to the one we made about explications of *boiling* in terms of forcible expulsion of water vapor.

The phenomenon I am pointing out in Barwise and Perry's scheme is just a special case of the issue of explicational imputation in general. The reason that the problem shows up with *attitude* explication in Barwise and Perry's scheme is simply that the propositional-attitude relationships (and certain other similarly operating ones) are the only ones they explicate: other relationships and properties are represented by atomic symbols (e.g., **clever** in (BP3)).

Creary's Scheme

Creary (1979) has proposed a neo-Fregean way of using logic to represent propositional attitudes, for the purposes of AI. The proposal is a development of one by McCarthy (1979). It is also reminiscent of the system of Church (1951, 1973, 1974). There are other proposals on similar neo-Fregean lines in the philosophical literature (e.g., Smith, 1981).

In the intended interpretation of Creary's system, terms can denote concepts as well as "ordinary" things. In particular, terms can denote proposition-like and description-like concepts, akin to Fregean senses (Geach & Black, 1952). By these means, Creary is able to provide distinct logical formulae corresponding to distinct ways of interpreting propositional-attitude reports, and is able to get around the classical inferential problems resulting from opacity. In particular, the system copes with arbitrary nesting of propositional attitudes in a principled way, though at the price of introducing infinite "ladders" of concepts: there are concepts extending to concepts, concepts extending to concepts that extend to concepts, and so on. (We say that a concept *extends to* its extension. We reserve the verb "denote" for talking about an item in a representation scheme denoting something in the domain of discourse.) Another possibly troublesome issue is that Creary's logic contains special binding operators that take it beyond the confines of first-order logic and require special axioms and/or rules of inference (not given in Creary [1979]) to ensure that appropriate inferences can be effected. Intuitively these operators are counterparts, on conceptual planes, to the usual quantifiers. In any case, Creary's scheme is one of the most promising approaches to propositional attitudes that can be found in the AI literature.

Creary's scheme is explicitly intended only for GEN/AG use, within AI programs. However, it is reasonable to suppose that Creary would allow the GEN use to include NL use, so that natural language sentences could be rendered as expressions in the scheme. In most of this section we will assume that NL/AG use is allowed. This is mainly for presentational convenience, and the observations can readily be divested of their natural-language trappings.

In this section we shall see that if a cognitive agent were to use Creary's scheme to represent the world, it would systematically impute probably incorrect conceptual structures to other agents. For the moment we will assume that some cognitive agent X is using Creary's scheme as a way of representing its world and as a means for rendering inputed English sentences. Suppose X receives the sentence

(S5) *Mike believes that Jim's wife is clever.*

There are a number of non-equivalent ways (S5) might be rendered in Creary's scheme, but the simplest is

(C5-a) **believe(mike, Clever(Wife(Jim))).**

The symbol **mike** denotes the person Mike. The symbol **Jim** denotes a particular concept of the person Jim, and, as far as one can gather from Creary

(1979), this concept is a standard concept of Jim. The person Jim himself is denoted by the symbol **jim.** The symbol **Wife** denotes a function that when applied to some concept c of some man delivers the (descriptional) concept d of *the wife of [that man as characterized by c] as such.* The "as such" indicates that d is a complex concept involving the notion "wife of" and the concept c. The symbol **Clever** denotes a function that when applied to some concept c of some person delivers the propositional concept of *[that person, as characterized by c] being clever,* as such. This is a complex concept involving the notion of cleverness. The **believe** predicate is applied to a person and to a propositional concept.

Thus the term **Wife(Jim)** denotes a complex, descriptional concept. If Jim has a wife, then that concept extends to her; however, the important point about the formula is that it attributes to Mike a belief cast in terms of the descriptional concept—the question of the extension of that concept is a side issue.

Another important Creary interpretation of (S5) is:

(C5-b) **(∃P) (concept-of(P,wife(jim)) ∧ believe(mike, Clever(P))).**

The symbol **wife** denotes the wife-of function. The expression **concept-of(P,wife(jim))** says that P is *some* concept or other of Jim's wife. Intuitively, therefore, (C5-b) says merely that Mike believes that some entity is clever, where the concept that Mike uses to refer to the entity is unspecified (or, at least, unspecified by (C5-b)) but, according to X, actually extends to Jim's wife.

Note that the intuition underlying the use of the **Clever** and **Wife** functions in (C5-a) and (C5-b) is that Mike's belief is in some sense couched in a direct way in terms of wife-ness (in (C5-a) only) and cleverness (in both formulae). There is an implicit, intuitive assumption that Mike, and any other cognitive agents, are able to conjure with cognitive items or structures (of some sort) that represent wife-ness and cleverness, and in particular are able to combine such cognitive entities to form more complex structures. These observations will be exploited in a moment.

Notice on the other hand that (C5-a) embodies no claim that *Mike* conjures with entities representing the functions denoted by the Creary symbols **Clever** and **Wife.** What Mike does have is representational items representing the *clever* predicate and the *wife-of* function themselves. The Clever and Wife functions are (so far at least) merely tools *the agent X* uses to "mentally discuss" Mike.

Now suppose that X inputs the sentence

(S6) *George believes that Mike believes that Jim's wife is clever.*

Following closely the lines of Creary (1979), two major Creary interpretations of (S6) are

(C6-a) **believe(george, Believe(Mike, Clever$(Wife$(Jim$))))**

(C6-b) **believe(george, (Exist P$)(Concept-of(P$, Wife(Jim))**
And Believe(Mike, Clever$(P$)))).

Intuitively, (C6-a) says that George has in his belief space something that says what (C5-a) says. Similarly, (C6-b) says that George has in his belief space something that says what (C5-b) says. In these formulae we see terms that denote second-order concepts. For instance, **Jim$** denotes a concept of a concept of Jim, though Creary (1979) is not entirely clear on the nature of such second-order concepts. The term **Wife$(Jim$)** denotes a concept of the concept denoted by **Wife(Jim)** (just as **Wife(Jim)** denotes a concept of the entity denoted by **wife(jim)**). **Clever$(Wife$(Jim$))** denotes a concept of the propositional concept denoted by **Clever(Wife(Jim))**. Observe that **Believe** denotes the function that, when applied to a concept of a person and a concept of a propositional concept, delivers the propositional concept of that person (so characterized) believing that proposition (so characterized). The symbol **P$** is a variable. The symbol **Exist** is a special binding operator that is used in terms denoting senses that involve the notion of existential quantification.

Our point hinges on the nature of the concepts denoted by the second-order concept terms in (C6-a) and (C6-b). First, it is reasonable to suppose that **Wife$** is to **Wife** as **Wife** is to **wife**, and similarly for **Clever$**. That is, **Wife$** denotes the function that takes a concept c of a concept d (of a person) and delivers the concept of the *Wife-concept of d [as characterized by c] as such.* (Note the analogy with the earlier explanation of the symbol **Wife**.) The delivered concept, therefore, explicitly involves the *concept-construction* function that is denoted by **Wife**. Similarly, **Clever$** denotes the function that takes a concept c of a concept d (of a person) and delivers the concept of the *Clever-concept of d [. . .]* as such; the delivered concept explicitly involves the concept-construction function that is denoted by **Clever**.

The trouble then arising is that we must conclude that, intuitively, (C6-a) conveys that *George's belief is couched in terms of the* **Clever** *and* **Wife** *concept-construction functions* (as well as the relation of believing). This is analogous to (C5-a) conveying that Mike's belief is couched in terms of the wife-of function and cleverness property. It may help also to compare (C6-a) to

(C) **believe(george, Taller(Mike, Father(Mother(Pat)))).**

In this formula it is clear that George's belief is couched in terms of the father and mother functions (as well as the taller-than relation); but **Taller, Father,** and **Mother** are analogous here to **Believe, Clever$,** and **Wife$** in (C6-a). The point is that we now see that (C6-a) is a *deviant* interpretation of (S6), whereas Creary makes it out to be a *major* plausible possibility. (C6-a) is deviant because it takes George to be conjuring with concept-construction functions that no one except a theoretician (e.g., Creary) could normally be expected to conjure with.

There is a slighly different way of presenting the case that may throw some light on the matter. Suppose that some technical version of English contains words **Wife-concept, Jim-concept,** and **Clever-concept** that correspond exactly to the function symbols **Wife, Jim,** and **Clever** in Creary's system. Then (C6-a) is really one natural interpretation of

(S6′) *George believes that Mike believes the Clever-concept of the Wife-concept of the Jim-concept*

rather than of (S6). (Here we assume that (S6′) is interpreted in a structurally faithful way.) By similar reasoning one can show that the other interpretations of (S6) that are possible in Creary's system are all natural interpretations of English sentences involving **Clever-concept** and **Wife-concept** rather than of (S6). Thus, (S6) is left with no natural interpretation in Creary's scheme.

Much as in our discussion of Barwise and Perry's scheme, we do not really need to bring in special vocabulary. Instead of (S6′) we could have used the following less convenient sentence, under an interpretation that is in part structurally *un*faithful:

(S6″) *George believes that Mike believes the result of applying the function Creary denotes by* '**Clever**' *to the result of applying the function Creary denotes by* '**Wife**' *to the concept Creary denotes by* '**Jim**'.

Here we must take the phrases "the result of... '...'" and "the concept Creary..." as being descriptions merely for the speaker/hearer's convenience, and not as portraying the structure of George's belief.

It is convenient to sum up the phenomenon noted as being an imputation, to cognitive agents, of features of X's particular method of describing cognitive agents. X's method uses concept-construction functions like Clever and Wife, and X, probably incorrectly, imputes the use of these functions to other cognitive agents. This observation about Creary's concept-construction functions was implicit in Barnden (1983), but was not put into the explicit broader context being developed here.

Confusions such as the one between sentences (S6) and S6′) could have practical importance at a sophisticated level. Consider the following dialogue, in which PROG is an AI program that is not based on Creary's system, whereas X is an AI program that is. (Moreover, X can understand references to Creary's theoretical constructs.)

JILL: George does not believe that Mike believes the Clever-concept of the Wife-concept of the Jim-concept.

X: Oh...so George does not believe that Mike believes that Jim's wife is clever.

PROG: Look, X, that doesn't follow. For one thing, you're assuming that George knows about Creary's concept-construction functions.
X: Doesn't everybody?

If PROG has also been based on Creary's system, then we would have expected its response to X's first statement to be something like, "X, why do you always state the obvious?"

It is not necessarily wrong to build an AI program X that uses Creary's scheme. It may well be that formulae like (C6-a) and (C6-b) are "sufficiently correct" to be heuristically good enough for X to act appropriately in its environment (as long as the English that X accepts does not contain words like **Wife-concept**, at any rate).

Consider now an "objective" semantic approach that uses Creary-like renderings of sentences. That is, the approach would be giving an NL/OB use to a Creary-like representation scheme. Such a semantic proposal would embody a commitment to the (*psychological*) thesis that agents conjure with Creary-like concept-construction functions. Although the theorist may have intended these imputations, it is equally possible that he or she may not have done, just as we suspect that Barwise and Perry are not aware of the imputations arising in their system.

Creary is not to be criticized for problems arising in uses of his scheme other than those he explicitly countenances. As stated before, he explicitly discusses only GEN/AG uses. It should be clear that the imputation difficulties we have uncovered when his scheme is given an NL/AG use carry over naturally to the GEN/AG case. The move is similar to the NL-GEN shift we effected in Section 3.

Expression-Denoting Schemes

A representation scheme is "expression-denoting" (XD) or "syntactic" if it includes items ιhat themselves denote representational expressions. If the denoted expressions are allowed to be in the scheme itself, then the scheme is a *reflexive* XD scheme. Our attention will be confined for now to reflexive XD logic schemes (where, moreover, *all* denoted expressions are in the scheme itself). Such schemes have been suggested as a way of dealing with propositional attitudes, and also with the alethic modalities of necessity and possibility. (See, for instance, Burdick [1982], Konolige [1982, 1984], Perlis [1985], and Quine [1981].)

The basic nature of the approach to propositional attitudes by reflexive XD logic schemes can be illustrated as follows. Consider the sentence

(S7) *Mike believes that Mary is clever.*

This could be rendered as

(XD7) **B(mike, 'clever(mary)').**

The XD feature here is B's second argument, which is an individual constant symbol that denotes the expression inside the quotation marks. (At least, the symbol so denotes in the intended interpretation of the logic.) The sentence

(S8) *George believes that Mike believes that Mary is clever*

would typically be rendered as

(XD8) **B(Geoge, 'B(mike, 'clever(mary)')').**

Here we have nested quotation corresponding to the attitude nesting.

Reflexive XD schemes are not confined to using quotation to denote expressions. For one thing, expressions can be quantified over and returned as function values, just as any other sort of object in the domain of discourse can. However, our discussion will not need to go beyond quotation.

The sort of XD scheme we have used here explicates belief to some small but significant extent, and thereby leads to harmful imputations. The explication resides in the fact that we have had to cast the having of a belief as being in some relation to an expression in the logic. To see the resulting imputation phenomenon, note that formula (XD8) is the natural rendering of

(S8′) *George believes that Mike believes-quoted-expression 'clever(mary)'*

under a structurally faithful interpretation. Here, *believes-quoted-expression* is a concocted verb that is rendered directly as the predicate symbol **B** in the logic. Thus, rendering (S8) as (XD8) effectively imputes to George a belief cast in terms of expressions in our logic and the relation *believes-quoted-expression*. As before, we can avoid the introduction of special vocabulary if we want to.

On the assumption that George is an ordinary human being, the imputation is clearly unfortunate if the logic scheme is being given an OB/NL use. It could also lead to harmful consequences if the logic scheme has an AG/NL use, as can be seen by constructing imaginary dialogues on the lines of the one we used in the discussion above of Creary's system.

All this is not to say that there is not *some* way of exploiting expression-denotation in a logic for propositional attitudes that reduces or avoids imputation problems. In fact, we suggest below a possible basis for the design of such a logic. What we have shown in the present section is that a straightforward, standard way of using the idea of expression-denotation leads to harmful imputation.

5. TOWARDS THE REDUCTION OF HARMFUL IMPUTATION

In this section we look at various ways of designing representation schemes so as to reduce the harm done by imputation. First, it is helpful to draw some distinctions between different types of imputation.

Degrees of Harmfulness in Imputation, and Strategy 1

Recall the discussion of

(S) *Mike believes that the water is boiling*

in Section 3. We looked at an explicative rendering of it that results in the imputation to Mike of a belief couched in terms of bubbling, hotness, and white mist. By considering a hypothetical AI program making inferences about Mike, we argued that the imputation could be harmful in practice, because it could lead to the program incorrectly inferring that Mike will remove some pictures from the room. However, it is also possible that *most* of the time the imputation will *not* be harmful in practice—it may be that Mike *normally* infers that water is bubbling, and so on, when it is boiling. If so, the imputation will, in a heuristic sense, be a reasonable one, and may even be positively benign, heuristically, in that it effectively embodies a useful, plausible inference about Mike's state of mind.

This situation should be contrasted with what happens in the case of the rendering of (S) that uses the explication of boiling in terms of forcible expulsion of vapor. Here, it is much less plausible that Mike, if an ordinary person, would have beliefs about boiling that are couched in terms of such vapor expulsion. (And we could increase the implausibility of the example by considering instead an explication in terms of molecular energy and attraction.) Thus, it is probably *not* heuristically reasonable for an AI program to interpret (S) in such a way as to impute to Mike a belief couched in terms of vapor expulsion. Such an imputation would be positively harmful if the program knew that Mike always feels anxious when he thinks about water vapor (because it reminds him of his poor performance in physics classes at school). It would normally be wrong for the program to infer that Mike feels anxious, on the basis of (S). Of course, the situation would be different if the program had special information about Mike, to the effect, say, that Mike is always thinking of the physical nature of the phenomena he observes.

Thus, we can say that one sort of explication of boiling (in terms of bubbling and so on) is heuristically reasonable, while the other (in terms of vapor expulsion) is heuristically unreasonable, if not completely incorrect. Heuristically unreasonable also are the imputations arising in Section 4 from renderings of nested propositional-attitude reports that explicate the inner propositional attitude in terms of Barwise and Perry's B_r relation, in terms of Creary's concept-construction functions, or in terms of quoted logic expressions. The imputations are of arcane theoretical notions to ordinary human agents.

Another conclusion we came to was that imputations that are heuristically reasonable when emanating from an AG use of a representation scheme (by an AI program, for instance) are in all likelihood sheerly *incorrect* when emanating from an OB use of a scheme (in a formal objective semantics of natural language, for instance). Even the explication of boiling in terms of bubbling, etc., leads to incorrectness; from the point of view of objective semantics, it is not good enough for it to be merely *normal* for Mike to *infer* that water is bubbling, is very hot, and is giving off a white mist when he believes the water to be boiling. As regards explication of propositional attitudes in terms of arcane Barwise and Perry, Creary or expression-quoting machinery, the resulting imputations are yet more seriously incorrect.

It is reasonable to suggest the following strategy for ameliorating the imputation problem in the case of representation schemes intended for AG use:

> STRATEGY 1. *Make sure that all explications that the representation scheme induces are cognitively reasonable for human beings. In particular, do not explicate propositional attitudes in terms of arcane theoretical constructs.*

By "cognitively reasonable" I mean that the explication should be one that it is reasonable to suggest is actually normally employed by human beings, just as it is reasonable to say that Mike normally infers that water is bubbling, etc., when he believes it to be boiling. Thus, cognitively reasonable explications lead to heuristically reasonable imputations. It is necessary to emphasize that the strategy only makes sense in the case of AG use of schemes, since in OB use even the heuristically reasonable imputations are undesirable.

We shall not attack the first sentence in the statement of Strategy 1 in all its generality. Instead, we turn now to look briefly at ways of achieving the intent of the second sentence.

Pursuing Strategy 1—Modal Logic

Possible Non-Attitude Explications. As illustrated by the consideration of sentences (S) and (S′) in Section 1, explications lead a modal scheme into making imputations, at least if the scheme is used in the seemingly natural way suggested.

Explications like that of boiling as forcible expulsion of vapor can of course be avoided in a modal scheme. One way is to use cognitively reasonable explications instead, as suggested in Strategy 1. This approach is suitable if the scheme is given an AG use. Another way is to avoid explication in the first place. For example, the scheme could contain an **is-boiling** predicate, and this would be used in the rendering of sentence (S). This approach is suitable if the scheme is given either OB use or AG use. Naturally, a great number of predicate symbols would be required to avoid all explications. This is a disadvantage, but may not be too high a price to pay (especially in the OB case) if it gets rid of imputation problems.

We now turn our attention to the question of the explication of propositional attitudes themselves, to see whether modal schemes suffer from imputational problems akin to those discussed in Section 4.

Lack of Belief Explication. Suppose we give an AG use to a modal logic scheme that includes a belief operator B, taking an (agent-referring) term as its first argument and a formula as its second argument. The sentence

(S9) *Mike believes that Mary is clever*

could then be rendered as

(M9) **B(mike, clever(mary)).**

Similarly, the sentence

(S10) *George believes that Mike believes that Mary is clever*

could be rendered as

(M10) **B(George, B(mike, clever(mary))).**

It is fair to say that the scheme does not explicate Mike's possession of a belief at all. We have simply used a "monolithic" belief operator B (that is not to be thought of as denoting a relationship in the sense that a predicate symbol denotes a relationship). This contrasts with the Barwise and Perry case, the Creary case, and the expression-denoting case as portrayed in Section 4. In Barwise and Perry's scheme, having a belief is explicated in terms of being in a relationship to a situation-type; in Creary's scheme, having a belief is explicated in terms of being in a relationship to a conceptual structure built by means of concept-construction functions; and in the expression-denoting scheme discussed, having a belief is explicated in terms of being in a relationship to a logical expression. The amount of detail introduced by these explications is not great, but they are significant explications nevertheless.

Since the scheme does not explicate belief, then, it does not introduce the sort of imputation of arcane, cognitively unreasonable theoretical machinery that we objected to in Section 4. Of course, the imputations arising from non-attitude explications, such as the explication of boiling, are still with us. These imputations may not be too serious in themselves, if the explications are cognitively reasonable.

What is serious is that a modal scheme encounters imputation problems if it explicates attitudes in a way that we have not yet considered, as we will now see.

Mutual Explication of Propositional Attitudes. We have concentrated throughout on belief, in the expectation that other attitudes lead to similar problems and are susceptible to similar treatments. For this tactic to be appropriate in this discussion of modal logic, we would have to assume that *none* of the attitudes receive an explicative treatment. But from the point of view of theoretical economy it might be suggested that some attitudes are explicated in terms of others. For instance, suppose knowledge is explicated as justified true belief by the logic, where justified belief is represented by a modal operator JB. So, the sentence

(S11) *Mike knows that Mary is clever*

could be rendered as

(M11) $\textbf{clever(mary)} \wedge \textbf{JB(mike, clever(mary))}.$

But then the rendering of

(S12) *George believes that Mike knows that Mary is clever*

as

(M12) $\textbf{B(george, clever(mary)} \wedge \textbf{JB(mike, clever(mary)))}$

imputes to George a belief that is of conjunctive form and is cast in terms of justified belief. This imputation can be seen by observing that (M12) is also the natural rendering of

(S12′) *George believes that: Mary is clever and Mike justifiably-believes that Mary is clever.*

It is not at all clear that ordinary human agents normally have beliefs cast in terms of justifiable-belief.

Modal Schemes: Other Problems. A modal scheme can avoid imputations arising from non-attitude explications and from mutual explication of attitudes by using sufficiently many predicate symbols and attitude operators. However, this solution is not particularly satisfactory, and in any case modal schemes suffer from other difficulties. Consider the sentence:

(S13) *Mike believes that some house is red*

and the renderings

(M13a) $\textbf{B(mike, } \exists \textbf{h(is-house(h)} \wedge \textbf{red(h)))}$

(M13b) $\exists \textbf{h(B(mike, is-house(h)} \wedge \textbf{red(h)))}.$

The first of these formulae is standardly taken to convey the de-dicto or "inner" reading, whereby Mike believes that some house or other is red and he does not have any specific house in mind. The second formula is standardly taken to convey the de-re or "outer" reading whereby there is an actual house of which Mike believes that it is red. But there is a troublesome "middle" reading whereby Mike's belief is based on a specific-house-characterization that does not necessarily characterize an actual house. For example, Mike may believe that the house at 600 Park Street is red, although in fact there is no house at that address.[3] Modal logic as usually conceived cannot render this "middle" interpretation, because of the lack of an alternative to formulae (M13a) and (M13b).

There is a way of rendering the desired middle reading by allowing not only ordinary quantification but also substitutional quantification (see e.g., Belnap & Grover, 1973). The rendering is

(M13c) $E\mathbf{h(B(mike,\ is\text{-}house(h) \wedge red(h)))}$

where E is an existential substitutional-quantification symbol. The formula is true just when there exists some expression $\mathbf{H}$ such that substituting $\mathbf{H}$ for $\mathbf{h}$ in $\mathbf{B(Mike,\ is\text{-}house(h) \wedge red(h))}$ leads to a true, well-formed formula.[4] The approach, however, is not able to deal with sentences that explicitly refer to properties of characterizations themselves. For example, the sentence

(S14) *Mike believes that some house of which he has a hazy idea is red*[5]

cannot be dealt with by an ordinary modal logic even with the addition of substitutional quantification. Inserting the conjunct $\mathbf{hazy(h)}$ into the body of formula (M13c) is wrong, since it would make that formula claim that the *house* is hazy. A similar objection would hold against a proposal to express middle readings by means of quantification over "world-lines" (cf. Kraut, 1983; Saarinen, 1981).

It is also no good using a higher order logic in any simple-minded way. One suggestion might be

(M14) $\exists_1\mathbf{p(B(Mike,\ hazy(p) \wedge is\text{-}house(\iota h.p(h)) \wedge red(\iota h.p(h))))}$.

Here, $\exists_1$ stands for quantification over predicates of one argument, and $\iota\mathbf{h.p(h)}$ means "the $\mathbf{h}$ such that $\mathbf{p(h)}$." The formula is inadequate because the predicates quantified over in a higher order logic are merely predicates

[3] See Saarinen (1981) for an extended discussion of "middle" readings. See also Hellan (1981).

[4] It is not entirely clear what the "inner substitutional" rendering $\mathbf{B(mike,}\ E\mathbf{h(is\text{-}house(h) \wedge red(h)))}$ expresses in commonsense terms. The rendering is different in meaning from the outer substitutional rendering (M13c), since the inner one certainly does not convey that there is a specific-house-characterization in Mike's belief. A rough guess at the meaning of the inner substitutional rendering is: *Mike believes that there is some red house that he could identify,* with the "some. .identify" of course being given inner scope. This guess can be backed up by appeal to a possible-world semantics of the modal logic, as the rendering would be true (in a particular world) iff for each of Mike's doxastic alternatives A to that world there is a term H_A denoting a red house in world A. Observe that the inner substitutional rendering conveys something different from what the ordinary inner rendering conveys, since $E\mathbf{h(is\text{-}house(h) \wedge red(h))}$ can be false when $\exists\mathbf{h(is\text{-}house(h) \wedge red(h))}$ is true. This happens if all the red houses are distinct from the houses that are denotanda of ground terms.

[5] This example sentence is to be interpreted in the following way: Mike believes that some house is red, where his belief is built from a house-characterization that is hazy. Mike's belief itself is not claimed to include the proposition that his house-characterization is hazy.

in extension, that is, they are just sets of tuples whose last elements are truth values. Our example, on the other hand, demands quantification over descriptions of some sort.

Another type of sentence that poses problems for modal approaches is illustrated by *Mike believes Bill's favorite proposition* and *Mike believes everything Bill believes.* (See also Burge, 1979.)

Therefore, although the modal approach may be good from the point of view of its ability to avoid undesirable imputations, it has other, independent, disadvantages that encourage us to look at other lines of attack.

Pursuing Strategy 1—Maida and Shapiro's Scheme

The intensional semantic networks of Maida and Shapiro (1982) embody an approach to attitude reports that is based on unexplicated attitude operators, and therefore may share the advantage that modal schemes have of avoiding unwelcome imputations arising from the nesting of attitudes. However, the scheme suffers from difficulties concerning the exact theoretical viewpoint that is being taken. These difficulties are explained in Barnden (1985). Although they interact closely with the problems of imputation, they are distinct, and it is therefore inappropriate to discuss the scheme in detail here. It is not yet clear whether a correction of the problems in the scheme introduces unwelcome imputations.

The scheme as it stands does already lead to possibly unwelcome imputations. Consider the sentence

(S15) *Mike believes that Jim's mother is clever*

where the phrase *Jim's mother* is to be taken "de re" or "referentially"—that is, it is a person-characterization that is being used for the hearer's benefit, with no commitment as to how Mike is thinking of the person (Mike need not even know of Jim). The Maida and Shapiro network representation for (S15) involves a node, standing for the unknown person-idea that Mike is employing in his belief, that is linked by a co-reference relationship to a node standing for the descriptional intension "the mother of Jim." Trouble arises from this coreference relationship when we move to representing the idea that George believes that Mike believes that some person is clever, that person being thought of as Jim's mother by *George,* but not necessarily (in George's view) by Mike. The natural way to proceed in Maida and Shapiro's scheme is to take the (S15) node, as it stands, as the OBJECT argument in a belief proposition that has the George node as AGENT argument. But then the upshot is to impute to George a belief couched partly in terms of the coreference relationship. More exactly, George is held to believe that: Mike believes that some person is clever where Mike's concept of the person is coreferential with the the-mother-of-Jim concept. Thus George is held to be thinking about the the-mother-of-Jim concept and the co-referentiality relationship. Although it might be argued that this imputation is cognitively reasonable, *it is up to Maida and Shapiro to make such an argument.* And in any case, the imputation is arguably less cognitively reasonable than, say, imputing to George a belief cast partly in terms of the notion that Mike has a person-concept that *describes Jim's mother.*

Thus, although the Maida and Shapiro scheme does not lead to imputation by virtue of explicating the belief relationship itself, it does lead to imputations by virtue of explicating some aspects of particular types of belief in terms of co-referentiality.

The scheme as presented in Maida and Shapiro (1982), and also in later papers such as Rapaport and Shapiro (1984), appears to have no general provision for quantifying over intensions, or for making statements about intensions themselves rather than about the "real-world" extensions of the intensions. This is in spite of the fact that intensions (concepts) are what are denoted by network nodes. Therefore the scheme seems unable to deal with sentences like "Mike believes that some house he has a hazy idea of is red" and "Mike believes everything Bill believes" that, as we saw, pose problems for modal approaches. The former sentence requires the haziness property to be predicted of a house-intension, and the latter requires quantification over propositional intensions.

A Second Strategy—Packaging Explications

Strategy 1 follows the line that if an explication is cognitively unreasonable, it had better be abandoned. However, there is an alternative:

> STRATEGY 2. *If an explication is cognitively unreasonable, then contain its effects by packaging it tightly by means of an abstraction operator.*

The strategy is obscure thus stated, but will now be illustrated by means of a major hypothetical modification of Creary's approach.

We will suggest the nature of the approach by a series of examples involving the renderings of natural language sentences by an agent X. Consider first the sentence

(S16) *Mike believes that Mary is clever.*

Agent X can render this as

(P16) $\mathbf{B}(\mathbf{mike}, x_M(\$_M(\mathbf{clever}), \$_M(\mathbf{mary})))$.

The believing has been explicated in terms of the **B** predicate, the x_M function and the $\$_M$ function. The $\$_M$ function delivers Mike's standard concept of a thing. To get George's standard concepts of things we use the function $\$_G$; and so on. We also assume that there is a concept-construction function x_A for each agent A. This function is applied to a concept P of a property and a concept C of something, and delivers the agent A's concept of that thing having that property (the thing and the property being characterized by C and P, respectively). If the function is applied to concept R of a relationship and concepts C_i of some things, then it delivers A's concept of the things being in that relationship (as characterized by C_i and R, respectively).

Now consider a familiar sentence:

(S17) *Mike believes that the water is boiling.*

The complement of this is

(S18) *The water is boiling.*

Assume for simplicity that X explicates boiling as expulsion of vapor, and that X renders (S18) as

(P18) $\exists \mathbf{v}.\ \mathbf{vapor}\ (\mathbf{v}) \wedge \mathbf{expelling}(\mathbf{w},\mathbf{v}).$

The key to the new approach to sentence (S17) is that we refrain from using (P18) as it stands. Instead, we first package up the explication by converting that formula into

(P18λ) $[\lambda \mathbf{u}.\exists \mathbf{v}(\mathbf{vapor}(\mathbf{v}) \wedge \mathbf{expelling}(\mathbf{u},\mathbf{v}))]\mathbf{w}.$

We assume that Mike possesses some standard concept of the property denoted by the lambda expression, and some standard concept of the water. (The property is just the property of boiling, from X's point of view.) Then, X can render (S17) as

(P17λ) $\mathbf{B}(\mathbf{mike},\ x_M(\$_M[\lambda \mathbf{u}.\ \exists \mathbf{v}(\mathbf{vapor}(\mathbf{v}) \wedge \mathbf{expelling}(\mathbf{u},\mathbf{v}))],\ \$_M(\mathbf{w}))).$

The point is that there are no assumptions about the form of Mike's concept of the property of expelling vapor. Specifically, there is no assumption that this concept is a complex one that involves the concept of being vapor and the concept of expulsion. Thus, (P17λ) does not say that the concept which is Mike's belief, and which is delivered by the x_M application, involves the concept of being water vapor and the concept of expulsion. Consequently, the approach avoids the imputation that we were concerned about earlier.

Also, note that the sentence

(S17′) *Mike believes that the water is expelling water vapor*

is naturally rendered by X as follows, assuming a structurally faithful interpretation:

(P17′) $\mathbf{B}(\mathbf{mike},\ x_M^{\exists V}(x_M^{\wedge}(x_M(\$_M(\mathbf{vapor}),\ \mathbf{V})),$
$x_M(\$_M(\mathbf{expelling}),\ \$_M(\mathbf{w}),\ \mathbf{V})))).$

(Here we have introduced a conjunctive-concept construction operator $x_M^{\wedge}$ and a variable-binding existential-concept construction operator $x_M^{\exists}$. The variable ranges over concepts.) Thus the sentences (S17) and S17′) get different renderings, and are therefore not confused.

A sentence which, when suitably interpreted, does have the same rendering as (S17) is

(S17″) *Mike believes that w has the property denoted by*
'$(\lambda u.\ \exists v(vapor(v) \wedge expelling(u,v)))$'.

The required interpretation is one that takes the phrase *the property*. . . 'λ. . . ' *referentially,* and not with structural faithfulness. That is, this phrase is merely *our* characterization of the property, and we leave unspecified the nature of Mike's concept of it (except in so far as we assume that this concept is Mike's standard concept of the property).

So, it seems that we can avoid some imputations, just by packaging the relevant explications. Of course, we can use the same device to deal with cognitively reasonable explications like that of boiling as bubbling and being very hot. So the approach outlined seems to be a very promising one. We will have to temper this promise in a moment, however. Before doing so, we should note that the packaging trick does no good in a modal logic.

A Packaging Attempt in Modal Logic. Considering again the sentence (S17), we could try to render it modally by using the packaged version, (P18λ), of its complement. This would yield

(M17λ) $\mathbf{B(mike,\ [\lambda u.\ \exists v(vapor(v) \wedge expelling(u,\ v))]w)}$.

But this is the natural rendering, under a *structurally faithful* interpretation, of

(S17‴) *Mike believes that the water is-such-that it is expelling vapor.*

Thus, we still have the imputation to Mike of a belief cast in terms of vapor and expulsion. The reason that the packaging trick did not work is that the modal logic contains nothing corresponding to the $\$_M$ function. This function was used in (P17λ) to, so to speak, insulate Mike from the explication of the property denoted by the lambda expression.

Explications of Propositional Attitudes. So far, therefore, it looks as though the Creary-based approach being outlined has greater power to avoid imputations than modal logic has. But what about imputations possibly arising from the explication of belief, which is where modal logic performed well?

We noted that the approach explicates belief in terms of $\$_A$ and x_A functions. This explication can be packaged, just as an explication of boiling can. Thus, in rendering

(S19) *George believes that Mike believes that Mary is clever.*

we do not work from (P16) as it stands. Rather, we first package up the belief explication, getting

(P16λ) $[\lambda \mathbf{a,f,x}.\ \mathbf{B}(\mathbf{a},\ x_{\bullet}(\$_{\bullet}(\mathbf{f}),\ \$_{\bullet}(\mathbf{x})))](\mathbf{mike,\ clever,\ mary})$.

It is now straightforward to devise a rendering for (S19):

(P19λ) $\mathbf{B}(\mathbf{george},\ x_{G}(\$_{G}[\lambda \mathbf{a,f,x}.\ \mathbf{B}(\mathbf{a},\ x_{\bullet}(\$_{\bullet}(\mathbf{f}),\ \$_{\bullet}(\mathbf{x})))],$
$\$_{G}(\mathbf{mike}),\ \$_{G}(\mathbf{clever}),\ \$_{G}(\mathbf{mary})))$.

Here the term $\$_{G}((\lambda\ldots))$ denotes George's concept of a certain relationship, without introducing any commitment about the nature of that concept. In particular, the concept need not be built up from concepts of x and $\$$ functions. As a result, the rendering (P19λ) does not lead to an imputation to George of the use of such concepts. To put it another way, the following sentence has a structurally faithful rendering that is distinctly different from (P19λ):

(S19′) *George believes that Mike believes the propositional-concept he (Mike) constructs from his concept of being-clever and his concept of Mary.*

The structurally faithful rendering of this is the somewhat hair-raising formula

(P19a) $\mathbf{B}(\mathbf{george},\ x_{G}(\$_{G}(\mathbf{B}),\ \$_{G}(\mathbf{mike}),$
$x_{G}(\$_{G}(x_{M}),\ x_{G}(\$_{G}(\$_{M}),\ \$_{G}(\mathbf{clever})),\ x_{G}(\$_{G}(\$_{M}),\ \$_{G}(\mathbf{mary})))))$.

This is obtained from the unpackaged rendering (P16) of sentence (S16) using the same techniques whereby (P19λ) was obtained from (P16λ).

A sentence that does have the same rendering as (S19) is

(S19″) *George believes that Mike has the relationship denoted by (*λ. . .) to being *clever and Mary,*

as long as we take the phrase *the relationship. . . '(*λ. . .*)'* referentially (as for the corresponding phrase in (S17″)). The fact that this sentence does have the same rendering as (S19) shows that there is an imputation of some sort. Namely, X imputes to George the possession of *some* concept of the relationship X denotes by '(λ. . .)', the possession of a concept of the property of being clever, and the possession of a belief constructed from these two concepts. The second imputation appears cognitively reasonable in itself, and the third is reasonable if the first two are.

The first is not so clearly reasonable. Moreover, if instead of *Mary is clever* in (S19) we had a more complex phrase, the lambda expression would be more complex, and the imputation to George of a concept of the property it denotes becomes less plausible. Suppose that instead of *Mary is clever* we have *Mary is clever* and *Sally is tall.* The (unpackaged) rendering of *Mike believes that Mary is clever and Sally is tall* is

(U) $\mathbf{B}(\mathbf{mike},\ x^{\wedge}_{M}(x_{M}(\$_{M}(\mathbf{clever}),\$_{M}(\mathbf{mary})),\ x_{M}(\$_{M}(\mathbf{tall}),\$_{M}(\mathbf{sally}))))$.

If in the packaged version we want to ensure that all the instances of $\$$ and x go inside the lambda expression (while keeping *clever, tall, M, mary* and *sally* outside), this expression has to be

(λ) $\mathbf{\lambda a,f1,x1,f2,x2.}\ x^{\wedge}_{a}(x_{a}(\$_{a}(\mathbf{f1}), \$_{a}(\mathbf{x1})), x_{a}(\$_{a}(\mathbf{f2}), \$_{a}(\mathbf{x2})))$.

Similar observations apply to other possible replacements for *Mary is clever* in the example. Thus, we end up imputing to George an indefinitely large collection of concepts that get more and more arcane as the relationships they are concepts of get more complex.

Thus the packaging device does not get us very far when we apply it to the explication of belief in terms of the $\$$ and x functions. The approach is therefore inferior to the modal approach in regard to imputations arising from belief-explication, despite the fact that it is superior with regard to other sorts of imputation, as we saw earlier. On the other hand, the packaging approach does get over the problem the modal approach faced of imputations arising from explications such as that of knowledge as justified true belief. Consider sentence (S11) again, and assume its unpackaged explicative rendering is

(P11) $\mathbf{clever(mary)} \wedge \mathbf{JB(mike,}\ x_{M}(\$_{M}(\mathbf{clever}),\$_{M}(\mathbf{mary})))$

so that its packaged rendering is

(P11λ) $[\mathbf{\lambda a,f,x.\ f(x)} \wedge \mathbf{JB(a,}\ x_{a}(\$_{a}(\mathbf{f}),\$_{a}(\mathbf{x})))](\mathbf{mike, clever, mary})$.

Then, in constructing a rendering of (S12) we apply $\$_{G}$ to the lambda expression here, and thus protect George from having a conjunctive belief couched in terms of justified-belief.

Before going on to comment on how the present approach's inferiority to modal logic in respect of belief-explication can be diminished, we should comment on the fact that we have unabashedly adopted the use of higher order constructs. We have even allowed $\$$ functions to be applied to $\$$ functions. Moreover, we have introduced the special variable-binding operators x^{3}. These moves promise considerable extra complication for any attempt to put the approach on a firm technical footing.

Back to Strategy 1—A Creary-Based Proposal

In the previous subsection, we proceeded on the implicit assumption that the explication of belief in terms of the x and $\$$ functions *was* to be packaged and therefore prevented from giving rise to imputations. However, there is a case for the idea that in fact such imputations would be cognitively reasonable. Thus, perhaps formula (P19a) is after all a heuristically adequate rendering of sentence (S19). As stated before, rendering the sentence thus is tantamount to confusing it with sentence (S19′), and therefore to imputing George with a belief that features the x and $\$$ functions. On the other hand, we would avoid the unreasonable imputations that we were faced with before (recall sentence (S19″)). At the same time, we would still have the ability to effect advantageous sorts of packaging. For instance, formula (P17λ) could still be our rendering of sentence (S17), and we could maintain a packaging of the explication of knowledge in terms of true justified belief.

I do not wish to make a final judgment here about whether the present proposal's imputations (of beliefs couched in terms of the x and $ functions) are really cognitively reasonable or not. I would certainly claim that they are *more* reasonable than the imputations pointed out in Section 4. That is, it seems to me more accurate to claim that beliefs are cast in terms of the x and $ functions than that they are cast in terms of the Barwise and Perry mental-classification (B_r) relationship, the Creary concept-construction functions, or quotations of internal formulae.[6]

Earlier in this section I mentioned certain disadvantages of modal schemes. These disadvantages are overcome by Creary's scheme and by the proposal in this section, although I do not go into the matter here.

An Expression-Denoting Variant. In a Creary-like spirit, we have taken various terms in the proposed logic to denote *concepts*. However, we could equally well take those terms to denote *representational expressions*, using a notion of representational expression that can be as broad as we please. For example, we can take the term $\mathbf{\$}_M(\mathbf{mary})$ to denote some (unspecified) collection of internal representational expressions, internalized natural language sentences, images, or whatever, that we consider Mike to deploy as his default way of thinking about Mary. Thus, one variant of our proposal is an *expression-denoting* proposal, though no denoted expression is ever *displayed* in the sense that expressions are displayed in *quotational,* expression-denoting schemes.[7]

An advantage of considering the expression-denoting variant is that it may be felt to lead to imputations that are more cognitively reasonable. That is, it may be more reasonable to suppose George's beliefs to be cast in terms of *expression*-returning functions x and $ than in terms of *concept*-returning functions. We get away from the idea that George thinks in terms of the vague-to-George notion of concept as opposed to possibly-less-vague-to-George notions such as those of internal natural language sentence, visual image, etc.

The move from concept-denoting terms to expression-denoting ones is similar in spirit to the point made by Burge (1979) that in taking a roughly

[6] The scheme being proposed is based on an early version presented in Barnden (1983). The approach to quantification in that paper, although appealing in that it avoided variables by being based on logical combinators, is undergoing modification in the light of the imputation issue. It should also be noted that it is technically preferable to use terms of form $\mathbf{\$}(\mathbf{a},\mathbf{x})$ and $x(\mathbf{a}, \mathbf{x},\mathbf{y},\ldots\mathbf{z})$ to terms of form $\mathbf{\$}_a(\mathbf{x})$ and $x_a(\mathbf{x},\mathbf{y},\ldots\mathbf{z})$, although the former lead to more complex formulae.

[7] The representational proposal in Barnden (1986a) is a development from the expression-denoting variant of the proposal made here.

Fregean approach one can replace the Fregean sense of an expression by the expression itself.

6. A THIRD STRATEGY—ELIMINATING SYNTACTIC COMPOSITIONALITY

So far we have been talking as if all the blame for imputation is to be laid at the door of explication or of the way explication is deployed. However, part of the blame should also be placed upon the "syntactic compositionality" of the representational approaches so far assumed.

The Nature of Compositionality

Roughly, "compositionality" in the sense used here is the technique of basing the formal rendering of a propositional attitude report on the formal rendering of the complement (the embedded sentence). More precisely, consider a sentence of form

y ATTs that C,

where *y* is an agent-reference expression, *ATTs* is some propositional attitude verb (such as *believes*), and *C* is the complement (e.g., *Mary is clever*). Then the rendering rule is *compositional* if the rendering R of *y ATTs that C* is syntactically built up from the formal rendering RC of C. This building-up-from can involve just embedding RC, as it stands, in a larger structure, or it can involve some simple syntactic manipulation before the embedding. The words "some" and "simple" here make the notion of compositionality fuzzy to some extent.

Consider for example the sentence

(S20) *Mike believes that Mary is clever.*

Suppose sentences are being formally rendered by means of a modal logic. Let the rendering of the complement, C, of (S20) be

(RC20.M) **clever(mary)**

and the rendering of the whole sentence be

(R20.M) **B(mike, clever(mary))**.

Then the sentence has been rendered compositionally in the strongest possible sense, since (RC20.M) has been plugged in without modification.

Now suppose sentences are being formally rendered by means of a quotational logic. Let the rendering of the complement, C, of (S20) again be

(RC20.Q) **clever(mary)**

and the rendering of the whole sentence be

(R20.Q) **B(mike, 'clever(mary)').**

Then the rendering is, again, strongly compositional, though not quite so strongly as before since (RC20.Q) has been subject to a slight modification before being plugged in. The modification consists merely of quoting the formula.

We get a significantly weaker, though still strong, form of compositionality in the case of the simplified Barwise and Perry scheme outlined in Section 4 (see Barwise and Perry's OB/NL Scheme). Take the complement rendering to be

(RC20.BP) **[clever(mary)]**

and the rendering of the whole sentence to be

(R20.BP) **[B$_r$(Mike, [clever(x)]);**
of(x,Mary)].

Then (RC20.BP) has been modified more dramatically than in the previous examples. Specifically, the formula has been subjected to an abstraction step. Nevertheless, it is reasonable to say that we have compositionality in a strong form still.

This strong compositionality applies generally to Barwise and Perry's renderings of propositional attitude reports, as the examples in Section 4 suggest. Actually, the formal renderings in the full treatment in Barwise and Perry (1983) are more complicated than the ones in Section 4, as they involve a provision for contextual factors. However, the essential compositionality as portrayed above still obtains.

Consider now rendering sentence (S20) in Creary's scheme. Assume the complement is rendered (yet again) as

(RC20.C) **clever(mary)**

and that the whole sentence is rendered as

(R20.C) **believe(mike, Clever(Mary)).**

We can feel happy in saying the rendering is still compositional in a fairly strong sense, since the modification to which the complement rendering has been subjected is still simple and straightforward. We call the modification a "lifting" of it up by one intensional level. Notice also that one standard rendering of

(S21) *George believes that Mike believes that Mary is clever*

is

(R21.C) **believe(george, Believe(Mike, Clever$(Mary$))).**

This is obtained compositionally from (R20.C) by lifting that formula up by one intensional level before embedding it.

Compositionality of renderings does not exclude the possibility that the complement rendering RC depends not only on the complement C but also to some extent on information picked up from other parts of the sentence (y or ATTs) or from contextual sources of some sort (as, indeed, happens in the full Barwise and Perry treatment). Such dependence could be of importance in the treatment of pronouns, indexicals, etc., within C, although these complications are not addressed in this paper. The greater the dependence of RC on y and ATT as opposed to C, the lesser the extent to which the rendering scheme should be regarded as being compositional.

The Third Strategy

Since we have assumed compositionality throughout, perhaps part or all of the blame for imputation should be laid at its door. This suggests the following strategy for avoiding imputations:

> STRATEGY 3. *Use non-compositional rendering techniques for propositional-attitude reports.*

To pursue such a strategy may strike some readers as heretical. Compositionality in the sense defined in this paper is currently such a common technique that its use seems to be regarded as requiring no justification.

The task of this paper is not to provide a particular approach to the rendering of propositional-attitude reports. Nevertheless, it is instructive to consider a hypothetical, highly oversimplified approach that is non-compositional. Consider an objective treatment of English specifying that the rendering of a sentence of form *y believes that P* is a formula that can be paraphrased in English as:

(T) *y's database contains the same internal representation as y would construct on inputing P,*

or more briefly: *y's database contains y's internal rendering of P.* Notice that the formula being paraphrased by (T) mentions the linguistic expression P itself.

Thus, part of the approach is to make assumptions about agents: for example, they have databases, and the databases can contain structures that we regard as the agents' renderings of English sentences. We do not make any assumptions about what those renderings are. The approach is non-compositional, because it does not rest on plugging in a rendering of P. Of course, the approach as described is highly oversimplified, as it takes no account of pronouns, indexicals and other context-dependent elements in P, and ignores the point that y may not render P in a unique way. As it stands it is also restricted to belief reports about agents that use natural language.

The idea of interpreting a propositional attitude report by appeal to a relation bringing in the linguistic complement itself is not new—for instance, Elgin (1985) suggests such a method, though she uses a less loaded relationship than our "y's database contains y's rendering of..." relationship, and y is taken to have the relationship merely to *some paraphrase* of the complement. The proposal in Brownstein (1982) is also relevant, as is the discussion of sentence-quotation approaches in Quine (1960).

Consider now the sentence:

(S22) *Mike believes that George believes that P*

(for some particular clause P). The rendering of (S22) is a structure R22 with English paraphrase:

(T22) *Mike's database contains Mike's rendering of "George believes that P".*

Here the quoted English expression (with P suitably replaced) is, as it stands, a component of R22. The point is that the semantic analysis of (S22) stops at (T22)—no analysis is performed of *George believes that P.* This "abandonment" of analysis may appear very strange at first sight. For one thing, we may feel that from (S22) we should be able to make deductions (or plausible inferences) about what Mike believes, based in some way on the meaning of *George believes that P.* Granted, we may well want to be able to make such inferences. But our *semantic approach* is not necessarily obliged to provide enough information on which to base them.[8] What is needed is an *additional, psychological* theory of the nature of Mike's mental processes. In particular, the theory should be about *his* sentence-rendering processes. The theory therefore does not have to be part of the objective semantics. By way of analogy, consider the sentence *Halley's comet is a hundred million miles from Earth.* We surely do not require that, in rendering this sentence, an objective semantics should itself go very far in helping us to deduce the comet's visibility from the Earth's surface. A whole body of extra theory is needed for that purpose. It just so happens that in the case of sentences about propositional attitudes, rather than about comets, the situation is obscured by the fact that what the sentences talk about (agents' renderings of sentences) is similar to what we are talking about (objective renderings of sentences).

One way of looking at the semantic approach under discussion is that it is *agent-theory parametrized.* That is, it is in some sense parametrized by par-

[8] Actually, Elgin's method (referred to above) does include in the rendering of a propositional-attitude report a truth condition derived from the complement. This truth condition is not stated to have any relationship to the agent holding the attitude; rather, the truth condition allows statements to be made about what follows from what the agent believes (desires, hopes, ...), but without an assumption that the agent therefore believes (desires, ...) them.

ticular psychological theories we might *also* have about the mental nature of agents. On the other hand, Barwise and Perry's objective semantics inadvertently plugs in a particular psychological theory about agents (to wit: that they can think in terms of B_r-classification, etc.). This plugging in is, of course, just the imputation phenomenon.

The Third Strategy in the AG/NL Case

Now consider an agent-based rendering approach on similar lines. That is, X's internal rendering of a sentence of form *y believes that P* is a structure that can be paraphrased as *y's database contains y's internal rendering of P*. Suppose that X inputs the sentence (S22). X therefore constructs R22, with paraphrase (T22), as an internal structure. Recall that R22 mentions the sentence *George believes that P* with P suitably replaced. The important observation is that we do not deduce from this that X constructs an internal structure with paraphase

(T22a) *Mike's database contains the statement: George's database contains George's rendering of P.*

This is because there is no necessity to suppose that X has any assumptions about the actual nature of Mike's internal renderings of inputed sentences, and in particular about the nature of Mike's renderings of belief reports. That is, it may be that all that X assumes is that Mike does construct renderings. For X to have assumptions about the nature of these renderings would be a distinct extra; and they could plausibly be regarded as extrapolations from some insight by X into his/her/its own internal mode of operation. Moreover, even if X did have the assumptions, a significant inferential step would be required to go from (T22) to (T22a).

If X did construct (T22a), X would be imputing to Mike X's own mode of operation. Further, if we assume that X does construct (T22a) (either instead of (T22) or as a result of inferring from (T22)) then the fact that X imputes something to Mike constitutes a "meta imputation" we make to X. It may, of course, be correct for X to infer (T22a), because it may be that our theory has it that Mike does in fact deal with belief reports in just the way X does. That in no way negates the existence of the mentioned imputations—it just makes them benign. It should be remembered, though, that our theory need not postulate that all agents deal with sentences in the same way: And if they do not, then the imputations could be harmful.

An upshot of all this is that our hypothetical rendering method makes a specific claim about the intrinsic nature of an agent X's internal renderings of belief reports, without imputing to X any particular view on the intrinsic nature of agents' renderings of belief reports (or of any sentences). That is, the rendering method does not impute itself to X. Instead, the rendering method has X explicitly referring to agents' rendering methods.

7. RELATIONSHIPS TO OTHER WORK

Various connections between our discussion and the literature on propositional attitudes should be mentioned. Taylor (1982, p. 192) points out that attempts to specify natural-language semantics can run the danger of ascribing (i.e., imputing) theoretical notions to ordinary speakers. Schwarz (1981) points out that certain notions, such as that of causal chains underlying reference, play a proper part in *agents'* views of other agents' beliefs and belief reports, but should not play a part in an *objective* view of belief. This claim is similar to our point that certain notions, such as the notion of the concept-construction functions x in Section 5 (see Back to Strategy 1—A Creary-Based Proposal) might play a part in the representation *by an agent* of other agents' beliefs, but may be suspect in a representation scheme with an objective use. Partee (1982) notes the problems that "theory-laden terms" such as the word "semantics" raise when they appear in the complements of propositional attitude reports. The problems hinge, effectively, on the different possible explications that such terms are susceptible to. It is possible that "believes" should be viewed as a theory-laden term in Partee's sense, and that the imputational problems raised by that word in a propositional-attitude complement are similar to the problems Partee is pointing out. My point, that an agent's NL representation scheme can generate incorrect but heuristically reasonable imputations, echoes Green's (1985) statement that natural-language sentences are merely rough, heuristically reasonable indications of the nature of agents' thoughts.

The imputation problem is relevant in the debate concerning the relative merits of "formal semantics" and "psychological semantics" of natural language (see e.g., Green, 1985; Moore & Hendrix, 1982; Partee, 1982; Peters & Saarinen, 1982; Saarinen, 1982). The former corresponds to our notion of "objective" semantic approach, and the latter to our notion of "agent-based" semantic approach. In that we have seen that the objective approach faces more difficulty with imputation, we have a reason for thinking that the agent-based approach is ultimately better. The discussion of the non-compositional approach to objective semantics suggests that an attempt at an objective semantics should avoid incorporating a specific theory about the intrinsic nature of agents' mental representations; it should instead confine itself to *referring* to such representations and relating them to natural-language expressions. (We said that such a semantics would be "agent-theory parametrized.")

It has been argued that a scientifically mature psychology should eschew any reference to "beliefs" or other propositional attitudes, and that the notion of propositional attitudes is purely a part of "folk psychology" (see e.g., Stich, 1983). (See Double [1985] and Lycan [1981] for counter-arguments. Lycan calls the anti-belief position"doxastophobic".) I do not come down on either side of this issue in this paper, but there are a few things I should point out. First, suppose that doxastophobia is justified. Consider a mature psychology's way of describing the world and the agents in it as an OB representation scheme RS. Then, RS must have some way or other of accounting for those states of the world that, normally, we partially characterize by means of belief reports. The issues of explication and possible imputation are just as likely to arise for RS as they are for any of the schemes we have considered in this paper. (These schemes all involve some sort of predicate or operator corresponding directly to our chosen sense of the word "believes," whereas RS would use some elaborate substitute.) In particular, RS must account for world states we describe by means of nested belief reports; and it must moreover do so in such a way that whatever explication is used for the *inner* belief in a report does not lead to implausible imputations to people. Second, the doxastophobic position's being valid would not imply that people cannot be correctly described as manipulating internal representational expressions, for this is a separate issue from the question of whether the notion of belief makes scientific sense. But, assuming that people do manipulate representational expressions, it is also the case that the doxastophobic position does not force these expressions to eschew folk psychology. So, it would still be possible for people to use internal representational schemes (of which they may have no conscious idea) that are somewhat similar to the ones we have looked at. Hence, the imputation issue is raised with respect to *people's internal* representation schemes. And, naturally, the issue still arises with respect to artificial cognitive agents' representation schemes (which need pay no heed to the doxastophobic position even if in principle it is correct). Third, nothing I have said amounts to a claim that people *are* well described as manipulating representational expressions. (In fact, I suspect the claim to be true, but it is not part of the point of this paper.) Rather, I have merely claimed that *if* people are taken to manipulate representational expressions belonging to some scheme, then the possibility of implausible imputations arising from that scheme must be attended to. And, even if people cannot be decently described as manipulating representational expressions, they may still construe other people as manipulating representational expressions (whatever scientifically respectable construal is given to the word "construe" in this sentence).

There is clearly a close connection between the issue of imputation as we have described it and the issue of translation of sentences from one natural language to another, since our "renderings" are translations of a sort (albeit into a formal language). For instance, the concept of compositionality and the trouble it leads to arises in natural translation. I would claim that the *most accurate* translation of a sentence of form "y believes that P" into French is something on the lines of "y croit-en-anglais 'P'," where this French expression simply quotes the English complement. (Note the similarity to the non-compositional approach in Section 6.) Never mind that the

French sentence would not convey much information to a French person who knew no English—every language contains many sentences that convey no information to a hearer who lacks appropriate knowledge. Certainly, the compositionally derived sentence ‘y croit que Q,'' where Q is the French for P, may *usually* be a *more useful/comprehensible* translation in practice—but it is nevertheless merely a heuristic approximation to the accurate translation involving the quotation of P. In particular, by giving some English word within P a French *explication,* the translation ''y croit que Q'' can produce undesirable imputation of concepts to agent y. Thus, my position is opposed to that of Church (1954), who (in the context of translation into German) takes the Q-based translation to be the proper one. His example involves the translation of the word ''fortnight'' in P into a German explication tantamount to ''period of 2 weeks,'' and thereby makes a cognitively reasonable but possibly false imputation to the belief-holder.

In demonstrating imputations, we often used the strategy of showing that the putative rendering of a propositional-attitude report U1 is more naturally to be regarded as the rendering of another propositional-attitude report U2, where U2 is U1 with some term in the complement given an explication. We rested on claims that U1 and U2 say different things about the holder of the attitude. Such claims are therefore connected to the general question of the synonymy or otherwise of propositional-attitude reports with different complements. Some authors (e.g., Partee, 1982) have suggested that any change in the complement of an attitude report is likely to disturb the meaning of the report. Such failures of synonymy are crucial in the argument that Mates (1950) adduces against intensional isomorphism (a form of structural similarity) of attitude-report complements being sufficient for synonymy of the reports.

The imputation issue is intimately tied to broader, long-standing, and troublesome questions about the nature of concepts and of analyses of concepts. To appeal to an example used by Langford (1942) and Moore (1942), one can know that a certain object is a cube without knowing that it has 12 sides (cf. our examples about boiling), and this seems to show that having 12 sides should not be part of an analysis of cubeness. The question of the nature of conceptual analysis is in turn bound up with the ''paradox of analysis'' or ''Frege's Puzzle'' (Linsky, 1983).

We have looked only at a small selection of the representational proposals that have been put forward for dealing with propositional attitudes. There is no implication that only these proposals suffer from imputation problems. This paper sets the stage for imputational analysis of other schemes, such as those based directly on possible worlds (e.g., Moore, 1977; Nilsson, 1983) or on Meinongian theory (Castañeda, 1974; Rapaport, in press).

Although there is no space to argue the point here, the imputation problem is intimately connected with the different sorts of reading to which propositional-attitude reports can be subject ("de re", "de dicto", etc.). The connection between imputations and readings will be spelled out elsewhere (Barnden, 1986b).

CONCLUSION

The main purpose of this paper is to encourage a greater awareness of the issue of imputation on the part of theoreticians who design or deploy representation schemes. We found that in the OB case, all imputations, cognitively reasonable or not, should be carefully scrutinized, as they embody a commitment to a psychological theory—whereas the theoretician using the scheme may not have intended any such commitment.

In the AG case, cognitively unreasonable imputations should be avoided, but cognitively reasonable ones may be tolerated. Even when imputations are cognitively reasonable, however, they should be thought about carefully by the theoretician.

We looked at three alternative strategies for reducing the bad effects of imputation. Strategy 1 says that only cognitively reasonable explications should be used, and in particular that propositional attitudes should not be explicated in terms of arcane notions. The strategy can be followed in using modal logic in a straightforward way. Strategy 2 allows cognitively unreasonable explications, but says they should be "packaged" in such a way that they do not lead to bad imputations. The strategy was followed in the design of a concept-denoting proposal that allows the "default-concept" $ function to be applied to explicational lambda expressions. Strategy 3, suitable for the NL case, says that the rendering of propositional attitude reports should be non-compositional, and requires a formal representation scheme that includes the ability to quote natural language sentences (or structures very close to sentences).

The different strategies are based on different decisions about which of the factors contributing to imputation should be tackled. These factors are: explications, the way explications are deployed, and compositionality. Strategy 1 banishes cognitively unreasonable explications as such. Strategy 2 banishes simple-minded ways of deploying explications. Strategy 3 banishes compositionality.

The differences between the various uses to which a representation scheme can be put has a major effect in the choice of a representation scheme. For instance, in the AG case the Creary-based proposal in Section 5 may be adequate, if the imputation of beliefs couched in terms of the concept-returning or expression-returning functions x and $ is judged to be cognitively reasonable. But these imputations may be intolerable in the OB case.

Although it has been convenient to couch most of the discussion in the context of NL uses of schemes, non-NL GEN uses are also clearly important. We have indicated now and again how the NL considerations can be mapped over to the GEN case. The basic strategy is to say that a representational structure that could hypothetically be said to be an imputational rendering of a sentence is still imputational when it is not a sentence rendering (but is instead generated from other sources). That is not to say that a scheme that is found suitable for GEN, non-NL use is necessarily suitable for NL use, or vice versa. For instance, in the NL case (OB or AG) it may be adequate to render sentences non-compositionally in an internal representation scheme that allows terms to denote natural-language sentences. However, if we want a representation scheme to be subject to GEN use, with no orientation towards the rendering of natural-language sentences, then the internal representational expressions demanded by the non-compositional NL approach seem much more suspect than they do in the NL case.

Thus one broad lesson is that we must be very careful to distinguish between the various uses to which representation schemes can be put. Both the OB/AG distinction and the NL/GEN distinctions are important. Another broad lesson is that, since imputation problems often only show up when nested propositional attitudes are considered, we must be careful to attend to such nesting when designing representation schemes. We accordingly repudiate the strategy, adopted for instance by Levesque (1984), of deliberately avoiding the complex issue of nesting in an attempt to reduce problems. Naturally, we can agree that problem-decomposition is a good general heuristic; unfortunately, work saved by ignoring attitude-nesting tends to compound the problems to be faced when nesting is ultimately considered, so that the problem-decomposition attempt happens not to be beneficial.

It may have appeared from our discussion of AG/NL uses of representation schemes that there is an assumption that we want a scheme to capture the "correct" meaning of sentences. But there may be no objective criterion of correctness. (It may be misguided to think we can provide an objective semantics of natural language.) So, we should be careful to avoid making the mentioned assumption. What we really want in the AG/NL case is for agents to be able to distinguish between sentences that competent speakers would generally distinguish between (as to truth value), to be able to use incoming sentences to help internal processing to mirror the external world to a heuristically adequate degree, and to be able to produce sentences that appropriately guide (or deliberately mislead) other agents. Thus, the real content of our statements that certain representational approaches engender undesirable imputations to agents is that the approaches are failing to use and distinguish between sentences as appropriately as they might. Indeed, our talk of imputations should really be regarded as an intuitive way of conveying such inappropriatenesses.

The paper has not, of course, attempted to deal with natural-language phenomena in a fully realistic way. Natural-language issues have perforce been greatly over-simplified in order to allow clarification of the imputation issue. For instance, the important issues of pronouns and indexicals in attitude-report complements have been ignored. There is, though, no obvious reason to fear that the paper's considerations will break down if put into a more realistic context. Also, although our examples have mainly been to do with belief, the arguments can be straightforwardly extended to other attitudes.

I have deliberately avoided paying attention to the relationship of conscious attitudes to unconscious internal representations. The analysis of imputation will eventually have to take this matter into account, though. For one thing, there is the point that some things that cannot reasonably be imputed to a human being's conscious processes may be reasonably imputed to his/her unconscious mental processes. This is especially true if we hold that a conscious belief involves conscious internal speech on the part of the agent, because then we might want to argue that nothing should be imputed (at the conscious level) that cannot be readily expressed in the particular natural language in question, even if it is readily expressible in unconscious mental representations. The issue is further complicated by the other sorts of conscious representational devices used by human agents, such as visual images of situations. To make matters worse, we must take into account the extent to which ordinary human agents, when they talk about or think about other people's beliefs, think about them as being based on conscious representational phenomena such as internal speech and visual images. That is, much of the time we need to be concerned not so much with what conscious and unconscious phenomena *really* underlie beliefs, but rather about what phenomena are *thought by ordinary agents* to underlie beliefs. Of course, this "thought by" could involve both conscious and unconscious processes.

Although the lack of a proper discussion of consciousness is an omission, it is one that does not put us at a disadvantage compared to most other discussions of propositional attitudes. The question of consciousness is, understandably, hardly ever explicitly addressed in such discussions.

Finally, our whole discussion has been centered on logic-based representation schemes. I hope it is clear, however, that nothing in the problems pointed out depends crucially on logic as such. For one thing, schemes based on networks or frames are so similar to logic schemes that they immediately come under the purview of the paper. Less obviously, even representation schemes that involve "non-propositional" elements such as, for instance, visual images, are likely to be susceptible to imputation problems. Imputation is to do with *explication* and *compositionality* in general, not so much with the particular form they take in logic-based schemes.

REFERENCES

Barnden, J.A. (1983). Intensions as such: An outline. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. Karlsruhe, West Germany.

Barnden, J.A. (1985). *Representations of intensions, and representations as intensions, and* *propositional attitudes* (Tech. Rep. No. 172). Bloomington, IN: Computer Science Department, Indiana University.

Barnden, J.A. (1986a). Interpreting propositional-attitude reports: Towards greater freedom and control. *Proceedings of the European Conference on Artificial Intelligence* (ECAI-86), Brighton, England.

Barnden, J.A. (1986b). *Readings and imputations: The explication of a problem in propositional-attitude representation.* Manuscript in preparation for submission for publication.

Barwise, J., & Perry, J. (1983). *Situations and attitudes.* Cambridge, MA: MIT Press.

Barwise, J., & Perry, J. (1985). Report of interview with Barwise and Perry. *Linguistics and Philosophy, 8,* 105-161.

Belnap, N.D., & Grover, D.L. (1973). Quantifying in and out of quotes. In H. LeBlanc (Ed.), *Truth, syntax and modality.* London: North-Holland.

Brownstein, D. (1982). Hard-core extensionalism and the analysis of belief. *Nôus, 16,* 543-566.

Burdick, H. (1982). A logical form for the propositional attitudes. *Synthese, 52,* 185-230.

Burge, T. (1979). Frege and the hierarchy. *Synthese, 40,* 265-281.

Castañeda, H.-N. (1974). Thinking and the structure of the world. *Philosophia, 4,* 3-40.

Church, A. (1951). A formulation of the logic of sense and denotation. In P. Henle (Ed.), *Structure, method and meaning: Essays in honor of Henry M. Sheffer.* New York: Liberal Arts Press.

Church, A. (1954). Intensional isomorphism and identity of belief. *Philosophical Studies, 5,* 65-73.

Church, A. (1973). Outline of a revised formulation of the logic of sense and denotation (Part I). *Noûs, 7,* 24-33.

Church, A. (1974). Outline of a revised formulation of the logic of sense and denotation (Part II). *Noûs, 8,* 135-156.

Creary, L.G. (1979). Propositional attitudes: Fregean representation and simulative reasoning. *Proceedings of the Sixth International Joint Conference on Artificial Intelligence,* Tokyo.

Double, R. (1985). The case against the case against belief. *Mind, 44,* 420-430.

Elgin, C.Z. (1985). Translucent belief. *Journal of Philosophy, 82,* 74-91.

Fagin, R., & Halpern. Y.J. (1985). Belief, awareness, and limited reasoning: Preliminary report. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* Los Angeles.

Fodor, J.A. (1978). Propositional attitudes. *The Monist, 61,* 501-523.

Geach, P., & Black, M. (Eds.). (1952). *Translations from the philosophical writings of Gottlob Frege.* Oxford: Blackwell.

Green, K. (1985). Is a logic for belief sentences possible? *Philosophical Studies, 47,* 29-55.

Halpern, J.Y., & Moses, Y.O. (1985). A guide to the modal logics of knowledge and belief: Preliminary draft. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* Los Angeles.

Hellan, L. (1981). On semantic scope, In F. Heny (Ed.), *Ambiguities in intensional contexts.* Dordrecht: D. Reidel.

Hintikka, J. (1983). Situations, possible worlds, and attitudes. *Synthese, 54,* 153-162.

Konolige, K. (1982). A first-order formalisation of knowledge and action for a multi-agent planning system. In J.E. Hayes, D. Michie, & Y-H. Pao (Eds.), *Machine intelligence 10.* Chichester: Ellis Horwood.

Konolige, K. (1984). *A deduction model of belief and its logics* (Report No. STAN-CS-84-1022). Stanford, CA: Department of Computer Science, Stanford University.

Konolige, K. (1985). A computational theory of belief introspection. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* Los Angeles.

Kraut, R. (1983). There are no *de dicto* attitudes. *Synthese, 54,* 275-294.

Langford, C.H. (1942). The notion of analysis in Moore's philosophy. In P.A. Schilpp (Ed.), *The philosophy of G.E. Moore.* Chicago: Northwestern University.

Levesque, H.J. (1984). A logic of implicit and explicit belief. *Proceedings of the National Conference on Artificial Intelligence,* University of Texas at Austin.

Linsky, L. (1983). *Oblique contexts.* Chicago: University of Chicago Press.

Lycan, W.G. (1981). Towards a homuncular theory of believing. *Cognition and Brain Theory, 4,* 139-159.

Maida, A.S., & Shapiro, S.C. (1982). Intensional concepts in propositional semantic networks. *Cognitive Science, 6,* 291-330.

McCarthy, J. (1979). First order theories of individual concepts and propositions. In J.E. Hayes, D. Michie, & L.I. Mikulich (Eds.), *Machine intelligence 9.* Chichester: Ellis Horwood.

Mates, B. (1950). Synonymity. *University of California Publications in Philosophy, 25,* 201-226. (Reprinted in L. Linsky (Ed.), *Semantics and the philosophy of language.* Urbana: U. Illinois press, 1952.)

Moore, G.E. (1942). A reply to my critics. In P.A. Schilpp (Ed.), *The philosophy of G.E. Moore.* Chicago: Northwestern University.

Moore, R.C. (1977). Reasoning about knowledge and action. *Proceedings of the Fifth International Joint Conference on Artificial Intelligene,* MIT.

Moore, R.C., & Hendrix, G.G. (1982). Computational models of belief and the semantics of belief sentences. In S. Peters & E. Saarinen (Eds.), *Proceses, beliefs, and questions.* Dordrecht: Reidel.

Nilsson, M. (1983). A logical model of knowledge. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence,* Karlsruhe.

Partee, B.H. (1982). Belief sentences and the limits of semantics. In S. Peters & E. Saarinen (Eds.), *Processes, beliefs and questions.* Dordrecht: Reidel.

Perlis, D. (1985). Languages with self-reference I: Foundations. *Artificial Intelligence, 25,* 301-322.

Peters, S., & Saarinen, E. (Eds.). (1982). *Processes, beliefs, and questions.* Dordrecht: Reidel.

Quine, W.V.O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Quine, W.V.O. (1981). Intensions revisited. In W.V. Quine (Eds.), *Theories and things.* Cambridge, MA: Harvard U. Press.

Rapaport, W.J. (in press). Meinongian semantics and artificial intelligence. In P. Simmons (Ed.), *Essays on Meinong.* Munich: Philosophia Verlag.

Rapaport, W.J., & Shapiro, S.C. (1984). Quasi-indexical reference in propositional semantic networks. *Proceedings of the 10th International Conference on Computational Linguistics,* Stanford University.

Saarinen, E. (1981). Quantifier phrases are (at least) five ways ambiguous in intensional contexts. In F. Heny (Ed.), *Ambiguities in intensional contexts:* Dordrecht: D. Reidel.

Saarinen, E. (1982). How to Frege a Russell-Kaplan. *Noûs, 16,* 253-276.

Schank, R.C. (1973). Identification of conceptualizations underlying natural language. In R.C. Schank & K.M. Colby (Eds.), *Computer models of thought and language.* San Francisco: Freeman.

Schwarz, D. (1981). Reference and relational belief: On causality and the pragmatics of 'referring to' and 'believing about'. In F. Heny (Ed.), *Ambiguities in intensional contexts.* Dordrecht: D. Reidel.

Smith, D.W. (1981). The Ortcutt connection. In F. Heny (Ed.), *Ambiguities in intensional contexts.* Dordrecht: D. Reidel.

Stich, S. (1983). *From folk psychology to cognitive science: The case against belief.* Cambridge, MA: MIT Press.

Taylor, B. (1982). On the need for a meaning-theory in a theory of meaning. *Mind, 91,* 183-200.