COGNITIVE SCIENCE **15**, 133–171 (1991)

Belief Ascription, Metaphor, and Intensional Identification

AFZAL BALLIM
University of Geneva, Switzerland

YORICK WILKS AND JOHN BARNDEN
New Mexico State University

This article discusses the extension of *ViewGen*, an algorithm derived for belief ascription, to the areas of intensional object identification and metaphor. ViewGen represents the beliefs of agents as explicit, partitioned proposition sets known as environments. Environments are convenient, even essential, for addressing important pragmatic issues of reasoning. The article concentrates on showing that the transfer of information in metaphors, intensional object identification, and ordinary, nonmetaphorical belief ascription can all be seen as different manifestations of a single environment-amalgamation process. The article also briefly discusses the extension of ViewGen to speech-act processing and the addition of a heuristic-based, relevance-determination procedure, and justifies the partitioning approach to belief ascription.

1. INTRODUCTION

An AI system that takes part in discourse with other agents must be able to reason about the beliefs, intentions, desires, and other propositional attitudes[1] of those agents, and of agents referred to in the discourse. This is especially so in those common situations where the agents' beliefs differ from the system's. Thus, the question of how to represent and reason about propositional attitudes is central to the study of discourse.

Clearly, this question is really about the beliefs, and so forth, that the system *ascribes* to the agents, on the evidence presented by the discourse itself and by context and prior information, because persons have no direct

We are indebted to Gerald Gazdar, Brian Slator, Nigel Shadbolt, Dan Fass, Louise Guthrie, and Sylvia Candelaria de Ram for thoughtful discussions. The errors are, as always, our own.

Correspondence and requests for reprints should be sent to Yorick Wilks, Computing Research Laboratory, New Mexico State University, Box 30001, Las Cruces, NM 88003.

[1] We use the term "propositional attitude" to cover beliefs, intentions, and so on, without intending to imply any specific philosophical view, such as one in which a state of belief (say) is a relationship between an agent and a "proposition."

access to each other's mental states. We view the ascription problem as being a fundamental one. It has been the focus of our past work on propositional attitudes (Ballim, 1986, 1987, 1988a; Ballim & Wilks, in press; Barnden, 1983, 1986, 1987a, 1987b, 1988, 1989a, 1989b, 1990; Wilks & Ballim, 1987, 1988, 1989a, 1989b; Wilks & Bien, 1979, 1983). Ascriptional reasoning is profoundly dependent on the communicative context, general information that the system has about the world, and special information the system has about the agents at hand. Moreover, there are major pragmatic features of discourse, such as speech acts, metaphor, and the determination of the intensional entities in play in a discourse, that any system for ascribing beliefs to agents must address. We would go further, and assert that even the most apparently superficial aspects of natural language understanding depend on belief ascription: such as prepositional phrase attachment. Anyone hearing a sentence with the all-too-familiar structure:

He told his mother about the murder in the park.

will interpret it differently according to whether he believes that the speaker believes there was a murder in a park and whether the speaker believes that the hearer believes it too. The function of our basic program *ViewGen* is to create, or as we shall call it, *ascribe,* environments into which appropriate beliefs can be segregated so that parsing and reasoning can be done in that limited environment.

We have described the basic algorithm in ViewGen in the publications above, and we address basic parsing issues seen as belief phenomena elsewhere. Here, our purpose is simply to review the basic ascription mechanism and then show its extension to the pragmatic phenomena in the title of the article.

In interpreting an utterance by an agent, the system must ascribe a speech act to that agent; and doing this is a matter of ascribing specific intentions, beliefs, desires, expectations and so on to the agent. Thus, speech-act ascription is an important special case of ascriptional reasoning. That speech-act considerations make reasoning about propositional attitudes essential for the computational modeling of discourse has been established at least since the work of Perrault and his colleagues (e.g., Perrault & Allen, 1980). A major difference between that work and ours is that they took the content of belief environments to be already established, whereas our approach is based on the real-time computation of the contents of such belief environments.

As for metaphor, to consider it at all in a study of propositional attitudes might initially seem unmotivated or overly ambitious. However, we are among those who hold that metaphor is central, not peripheral, to language use, and indeed, cognition in general (for related positions see, e.g., Carbonell, 1982; Hobbs, 1983a, 1983b; Johnson, 1987; Lakoff, 1987; Lakoff &

Johnson, 1980). We believe, in particular, that metaphor is *inextricably* bound up with propositional attitude processing for three main reasons:

1. A key aspect of a metaphorical view of a topic is seeing it *as something else:* even in such simple, conventionalized cases as *John caught a cold,* where a cold is seen as a missile or other object. This, we suggest, is no more than a special case of an *agent's view* of a topic, in the sense of a set of beliefs.
2. Many, if not most, beliefs arising in ordinary discourse and reasoning are at least partly metaphorical in nature. Consider, for instance, the beliefs that *Terrorism is creeping across the globe, Sally's theory is threatened by the experiment,* and *Prussia invaded France in 1871,* all of which are, in a broad sense, metaphorical. As an example of the difficulties that such beliefs raise, notice that the last one cannot in general be adequately represented by any literal sense representation for *Prussia,* since it may be important to take into account exactly how the believer may be viewing the invasion:

    - as a matter of the army of Prussia doing something,
    - of the Prussian government doing something,
    - or of the Prussian people as a whole doing something, and so on.

    The simple predicate notations commonly used in belief research lead us to overlook such basic representational issues.
3. People commonly (if not universally) think of minds and mental functioning in highly metaphorical terms—for instance, as physical containers of ideas, beliefs, intentions, and so on—those contents themselves being viewed metaphorically as active or passive physical objects of some sort. Thus, in a sentence like, *Mike believes that George believes that P,* we confront the issue of possible metaphorical views Mike may hold of George's beliefs. This issue, which is an important special case of (2), is studied in Barnden (1989a, 1989b, 1990), but for reasons of space is not addressed here.

The similarity in (1) is the main topic of this article. Note also that Davidson (1978) said that metaphor "is simply false belief." Our aim could be said to show that this is correct, but in a surprising and computationally realizable way. Our previous work was based on the use of *explicit belief environments.* Each of these is a group of propositions, manipulable as an explicit entity in the system, and which can, in ways we shall show, be thought of as *nested* within other such entities. The relation of nesting or containment represents the intuitive notion of a believer (an outer proposition group) having beliefs about other entity (the inner group). Our belief environments are akin to the belief spaces and other types of cluster or partition discussed more recently by authors such as Fauconnier (1985) and Dinsmore (1987). We also share a general belief in the primacy of intensional representation with Shapiro and Rapaport (1986) and their SNePS system. However, SNePS does not have any natural ánalogue of partitions or nestings of belief sets (the boxes that appeared in diagrams in Rapaport, 1986, just being a notational convenience), and so lacks a crucial feature of what we propose.

Maida's (1984, 1986) work shared many of the current concerns here: Maida linked belief ascription to analogical reasoning, and his diagrammatic representations of nested beliefs were isomorphic to those of Wilks and Bien (1979) and Shadbolt (1983). Maida's concerns were with the problem of shared reasoning strategies between believers and how, for example, it could be established that a dialogue partner also used *modus ponens*. We argue, on the contrary, that this phenomenon is best handled by general default assumptions, as are the concrete contents of beliefs. No finite set of dialogue observations could ever establish conclusively that another believer was using *modus ponens*. That being so, concentration on such issues that are not susceptible to proof seems, to us, only to delay the central issue, which is how to infer heuristically the actual beliefs of other believers. Maida (1983) was also concerned with the very important, and we believe quite separable issue, of a heuristic rule for identifying intensional individuals under different descriptions. Konolige's (1983) work had strong similarities to that just noted; Konolige considered what he called views, for which he wrote, for example, $v = John, Sue, Kim$ to mean *John's view of Sue's view of Kim's beliefs*. But he had no effective construction for the content of such views. Rather, Konolige was concerned with giving an account of limited deduction in such views, an important process, but not relevant to issues of constructing individuals' views. Dinsmore (1987) was concerned with what he termed the "algebra of belief spaces" but, although the term is highly general, the focus of his attention was always, in fact, the notions of presupposition and counterfactuals, which are not notions we treat explicitly here, and his treatment of them may well be compatible with our own general approach.

Our work has been closer in spirit to that of Perrault and others (e.g., Cohen & Levesque, 1980; Perrault & Allen, 1980); though without their (then) commitment to the language of speech-act theory and, most importantly, without their key assumption that the partitions among beliefs are all present at the beginning of the speech-act reasoning procedures. Our work makes no such assumption: For us, nested beliefs are not merely accessed, but constructed and maintained in real time, a position we find both computationally and psychologically more plausible. The Gedankenexperiment here is to ask yourself if you aready know what Mr. Gorbachev believes the U.S. President believes about Colonel Qaddafi. Of course you can work it out, but how plausible is it that you already have precomputed such nested viewpoints, in advance of any such consideration?

In general, our work has been, since that of Wilks and Bien (1979, 1983), to construct a formalism and programs (some would not abstain from the word "theory" here, but that difference of taste need not detain us, or see Wilks, 1990) that capture the heuristic belief ascriptions that individuals actually perform in the process of understanding and participating in dialgoue: That is to say, concrete beliefs and not merely meta-beliefs about the reasoning architecture of others, activities we suspect are rarely, if ever, undertaken. Thus, we have been less concerned with the general expressive powers of particular notations and demonstrations of their adequacy (as has been the central feature of most work on propositional attitude representation) than with the *content* of belief ascription. We suspect that the procedures we offer here could be applied to a large range of representational systems already available in the field.

The plan of this article is as follows: Sections 2 to 3 describe ViewGen, our present belief ascription system based on explicit proposition groups known as *environments,* and present justifications for the use of explicit environments. Section 4 discusses two issues that are important both for belief ascription and reasoning in general. The first is the notion of relevance, which is essential to realistic processing; the second is the intensional identification of objects, which, among other things, has a strong bearing on determining relevant beliefs. Section 5 forms the core of the article: It explains some profound connections that we see between belief ascription and metaphor, and describes how our current system is being extended to embody these connections. Section 6 considers the bearing of these issues on the processing of speech acts. Section 7 contains a general discussion, and Section 8 is the conclusion.

2. ViewGen: THE BASIC BELIEF ENGINE

A computational model of belief ascription is described in detail elsewhere (Ballim, 1987; Ballim & Wilks, in press; Wilks & Ballim, 1987; Wilks & Bien, 1979, 1983) and is embodied in a program called ViewGen. The basic algorithm of this model uses the notion of default reasoning to ascribe beliefs to other agents unless there is evidence to prevent the ascription. Perrault (1987, 1990) and Cohen & Levesque (1985) also recently explored a belief and speech-act logic based on a single explicit default axiom. As our previous work showed the default ascription is basically correct, but the phenomena are more complex (see the following) than are normally captured by an axiomatic approach.

ViewGen's belief space is divided into a number of topic-specific partitions (topic environments). These environments can be thought of as a less permanent version of *frames* (Charniak, 1978; Minsky, 1975), or more suitably, in terms of Wilks (1977), as "pseudotexts." In effect, a pseudotext is a

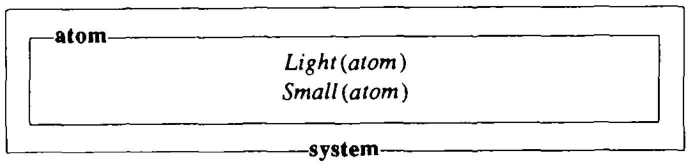

Figure 1. The system's view of an atom.

set of unsorted, unrefined items of knowledge. These pseudotexts are general items, and can be not only about individual objects, but also about abstract ideas and groups of things. The hierarchical and inheritance relations of pseudotexts are discussed in Wilks (1977) and Ballim & Wilks (in press). We jusify the general notion of explicit environment in the next section.

ViewGen also generates a type of environment known as a *viewpoint*. A viewpoint is some person's beliefs about a topic. Within ViewGen, all beliefs are ultimately beliefs held by the system (e.g., the system's beliefs about France, what the system believes John believes about cars, etc.) and so, trivially, lie within the system's viewpoint. The system's view of some topic (say, atoms) is pictorially represented as in Figure 1.

This diagram contains two types of environments: First, there is the box labeled with "system" at the bottom. This is a "believer environment" or "viewpoint." Viewpoints contain "topic environments," such as the box labeled with "atom" at the top of it. A topic environment contains a group of propositions about the "topic." So, for example, the diagram in Figure 1 conveys that the system believes that atoms are light and small. Topic boxes are motivated by concerns of limited reasoning (see Section 4.1 on relevance, and also Wilks & Bien, 1983). In short, it is envisaged that reasoning takes place "within" a topic environment, as if it were the environment of a procedure in a programming language.

Within ViewGen, environments are dynamically created and altered. ViewGen's "knowledge base" can be seen as one large viewpoint containing a large number of topic environments, with each topic environment containing a group of "beliefs" that the system holds about the topic. The reader should note that each proposition in a topic environment has at least one symbol identical to the name of the topic. Each such proposition is, therefore, *explicitly* about the topic. There are, however, implicit ways in which a proposition can be "about" (or "relevant to") a topic. The simplest cases are generated by inheritance in the usual way: For example, if John is a man, then any proposition in a "man" topic environment is implicitly or indirectly about John. However, we choose not to put such a proposition in the John topic box, and will justify that decision in Section 4.1 on relevance. Again, the same proposition can occur in more than one box, as would the expression asserting that an elephant was larger than an atom, for it is about both atoms and elephants, and should appear under both topics.

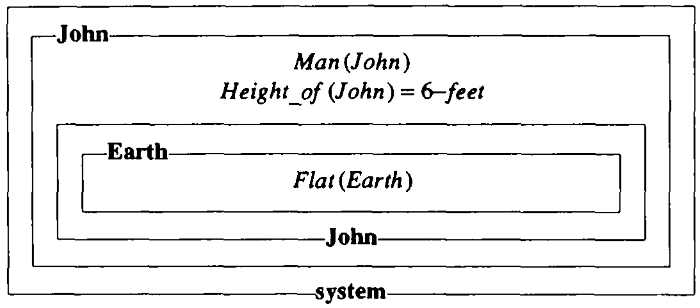

Figure 2. The organization of beliefs about and of John.

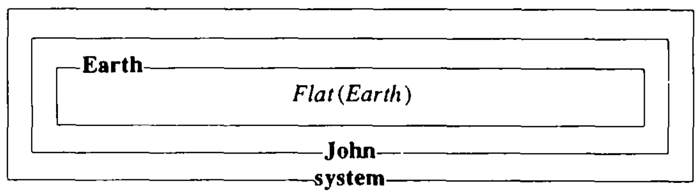

Figure 3. The system's view of John's view of the Earth.

If the topic of a topic environment is a person then the topic environment may contain, in addition to the beliefs about the person, a viewpoint environment containing particular beliefs held by that person about various topics. Normally, and for obvious reasons of efficiency, this is only done for those beliefs of a given person that are, as some would put it, reportable, which often means beliefs that confict with those of the system itself. For example, suppose the system had beliefs about a person called John who believes that the Earth is flat. (This is pictorically represented as in Figure 2.)

The John viewpoint, shown as the box with "John" on the lower edge, is a *nested viewpoint,* as it is enclosed within the system viewpoint shown (through an intervening topic environment about John, shown as the box with "John" on its upper edge). For simplicity, in the diagram of a nested viewpoint we often leave out propositions that are not in the innermost topic box: In this example, we would leave out the beliefs that John is a man, and that he is six feet tall. Further simplifying this, we often omit all but the innermost topic box, leaving only it and the viewpoint boxes. Hence, the diagram in Figure 2 would be simplified as in Figure 3.[2]

[2] We do not discuss here the issue of different mental descriptions under which John might have beliefs about the Earth. A case in which, say, John believes that a certain planet is flat, describing it mentally as the third planet from the Sun, can be handled in our system by having a complex topic-environment label, on the lines of the complex labels used on some occasions later in the article. Also, our techniques allow John to fail to realize that the third planet from the Sun is Earth (see 4.2.2).

The system stores its own beliefs, and the beliefs of other agents that differ from the system's own beliefs. Others' viewpoints are generated on demand a position we find both computationally and psychologically more plausible than the "prestored nesting" view mentioned in Section 1. The process of generating a viewpoint can be regarded as an *amalgamation* mechanism that ascribes beliefs from one viewpoint to another (or, "pushing one environment down into another"): ascribing certain beliefs, transforming some, and blocking the ascription of others.

The simplest form of this algorithm, described in Wilks and Bien (1979, 1983), is that a viewpoint should be generated using a default rule for ascription of beliefs. The default ascriptional rule is to assume that another person's view is the same as one's own *except where there is explicit evidence to the contrary*. An important special case of such examples is when the topic is the same as the agent, and we can illustrate with that. Suppose that at a certain stage in dialogue, the system, acting as a medical diagnostician, has the view that John is not healthy, and is six feet tall, although John believes himself to be healthy. This basic situation is represented pictorially in Figure 4. The more complex environment for the system's view of John's view of himself can be generated by trying to ascribe the beliefs from the system's topic enviornment about John to the topic environment about John within John's viewpoint (where, as always, the last expression must be glossed as "the system's view of.."). One of the two beliefs survives the attempt but the other is blocked, giving the state in Figure 5. This can be pictured in the simplified (or as we shall call it, *compressed*) manner as in Figure 6.

We see that in examples of this sort, where the topic is also the agent into whose environment as ascription is being attempted, propositions in an outer topic environment *E* are *pushed inwards* into a topic environment (for the same topic) within a viewpoint nested within *E*.[3] Such inward pushing is central to our later observations of intensional identification and metaphor.

The example just outlined demonstrates the basic ascription algorithm and a simple case of ascriptions being blocked. However, belief ascription is a far more complex phenomenon, and the key to our method is the delimitation and treatment of cases where the default algorithm is incorrect. But even the default algorithm requires, for its operation, a notion of blocking beyond that of explicit contradiction: For example, the proposition *Healthy(John)* should be able to block *Sick(John),* if *Sick* and *Healthy* are

[3] In our example we do take John to be having beliefs which he recognizes as being about *himself*. There are also unusual cases in which it is appropriate to take John's concept of himself to differ from that concept of his that most closely corresponds to the system's concept of him. (For example, he may be an amnesiac who has forgotten who he is, but nevertheless has beliefs involving a person that the system would say was he; cf. Rapaport, 1986). Such cases can easily be handled within the approach to intensional objects in Section 4.2.2, below, basically by having two environments on the topic of John within John's viewpoint.

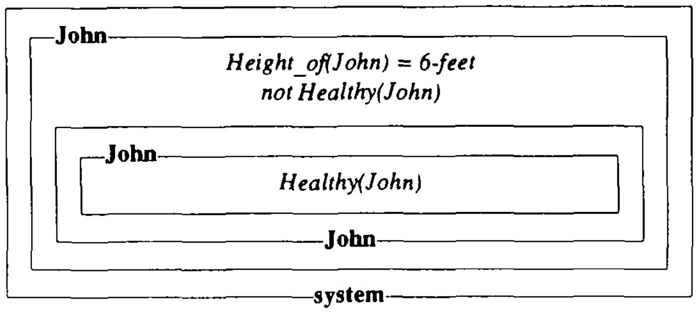

Figure 4. Beliefs pertinent to John.

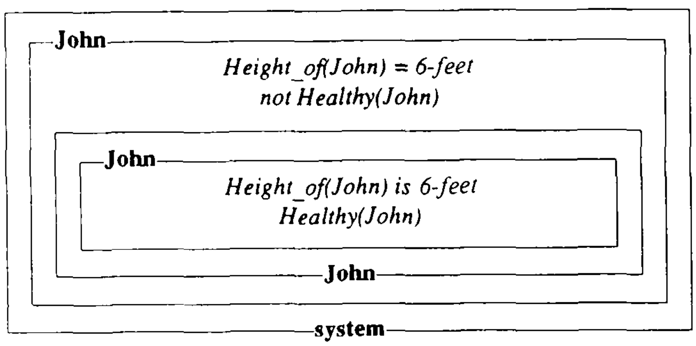

Figure 5. Generating John's beliefs about himself.

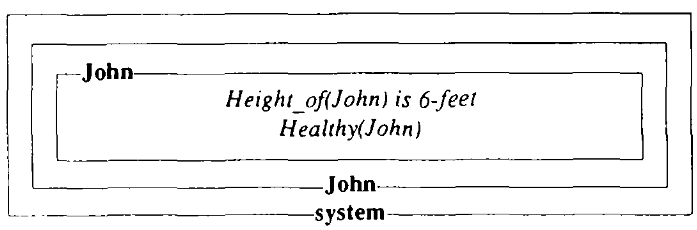

Figure 6. Simplified form of Figure 5.

known to be incompatible predicates. Similarly, we appeal later to blocking that arises from incompatible function values, as in the blocking of *Eye-color(Frank) = Green* by *Eye-color(Frank) = Blue*. The more significant complication is that there is an entire class of beliefs that require the opposite of the default-ascription rule given above. We call these *atypical beliefs* and they include technical expertise, self-knowledge (itself a form of expertise), and secrets. For example, beliefs that I have about myself, such as how many fillings I have in my teeth, are beliefs that I would not normally ascribe to someone else *unless I had reason to do so* (if, say, the person, to whom I was ascribing the belief was my dentist). A representation based on lambda expressions is used in dealing with atypical beliefs, and is described elsewhere (Ballim, 1987; Ballim & Wilks, in press; Wilks & Ballim, 1987), and follows a suggestion originally made by McCarthy and Hayes (1969). This combination of a basic default ascription rule, augmented by a mechanism for dealing with atypical belief, is an original algorithm and has not, to our knowledge, been described or tested elsewhere in the literature.

The essential feature of this notation is that lambda expressions, as in the following example

$$(\exists X).\{X = (\lambda y(Cure-for\ y)tuberculosis)\}$$

can only be evaluated by qualified believers (e.g., physicians or informed lay people in this case) in appropriate environments. Yet, anyone can believe the Fregean triviality expressed by the above sentence when it is unevaluated (and it is vital that they can) but a nontrivial interpretation can only be placed on it by those who can evaluate the lambda expression in an environment. In a crude sense therefore, the lambda expressions correspond to intensional representations, and their evaluations correspond, when available, to extensions, or at least other intensions in those situations where the evaluation of such an expression produces yet another lambda expression (see also Maida, 1983).

The above expressions, for example, might evaluate to another lambda expression using a predicate *Sulfonamide-drug,* for whose evaluation a particular drug might be appropriate. Each evaluation would require an environment whose ''holder'' was qualified to perform it. It is really this possibility of successive evaluations of expressions that justifies the abstraction capacity of the lambda notation, because it could well result in expressions, such as a conjunction of predicates, for which there is no single predicate that deals with the problem of the over-application of the main default rule of ascription, since the ascription of unevaluable expressions, about, say, the number of my own teeth, to you does not lead to undesirable results.

It should be noted that in blocking the ascription of a proposition, from one environment to another, we often need to consider not just whether it contradicts a proposition in the target environment, but also whether some combination of propositions in the source environment contradict some propositions in the target environment. This is considered in more detail in Ballim and Wilks (in press).

While on the subject of ascription blocking, we should mention that, in principle, a proposition *P* should not be ascribed from an environment *E1* to an environment *E2* if some presuppositions used in deriving *P* are blocked from being ascribed to *E2*. Thus, in principle, the issues addressed by truth-maintenance systems arise for us, although they are not yet addressed by ViewGen.

3. WHY EXPLICIT ENVIRONMENTS?

In a realistic discourse, the system has to make rapid decisions about the sets of propositions believed by the agents. Now, ascription can involve a significant amount of work in modifying an existing proposition before ascribing it, or in checking that there is no contrary proposition blocking the ascription (Ballim & Wilks, in press; Wilks & Ballim, 1987). Therefore, it is beneficial to minimize the number of propositions ascribed (as long as the techniques for minimization do not themselves take up too much time). One technique for limiting the ascription is to ascribe only those propositions that are deemed relevant according to some set of efficient relevance-determination heuristics (see Section 4.1).

Suppose the system has already constructed its own topic environment *R,* containing system beliefs about Reagan. The "default-ascription rule" used in ViewGen to construct or expand John's topic environment *JR* concerning Reagan is then just to push propositions in *R* down into *JR*. The pushing of a proposition may be blocked, because, for instance, it is explicitly contradicted by a proposition in *JR,* or because it is political expertise which should not be ascribed to the politically inexpert John. Also, propositions may need to be modified rather than blocked (Wilks & Ballim, 1987). Therefore, the pushing process as applied to *R* does require separate processing of individual propositions in *R*. However, the explicitness of *R* as a *group* is nevertheless important because *R* is likely to be the result of a significant amount of *knowledge-intensive, relevance-determination work* (see Section 4.1). This work may have involved the processing of system beliefs that are not about Reagan in any directly obvious, explicit way. Once the system has created *R* for the purposes of its own reasoning about Reagan, *R* is immediately available to help in constructing environments such as *JR,* for the purposes of the system's reasoning about various other agents' reasoning about Reagan. If beliefs were not parceled up in explicit environments, the ascription beliefs about Reagan to those agents would be likely to involve essentially *duplicated* relevance-determination work similar to what is necessary to create *R*. In sum, one justification for environments—proposition groups that are explicit in the above sense—is that they serve to reduce the amount of work dictated by considerations of relevance.

Also, the pushing down of system beliefs about Reagan into John's viewpoint could involve the conjoint examination of several such beliefs, rather than examination of them one at a time. It makes it especially important for the system to be able to determine *quickly* which of its beliefs are relevant to Reagan. A similar observation holds for pushing of beliefs at deeper levels of nesting, as in the attempted pushing down of John's beliefs about Reagan into a Bill viewpoint nested within John's.

People talk explicitly or implicitly about *sets* of beliefs (and other propositional attitudes) held by agents. For instance, someone might say "John's beliefs about New Mexico are confused." This sentence is best interpreted as conveying that John's beliefs are, as a set, inconsistent in some sense, rather than,as conveying something about individual beliefs of John. Explicit topic environments and viewpoints give us a handle on dealing with such cases.

Work by other researchers tends to support the importance of explicit environments. Fauconnier's (1985) mental space theory used environment-like entities to explore a number of the same issues as in this and previous articles, from a linguisitic perspective. Although Fauconnier's account was not procedural in nature, there are certainly analogies between our default-ascription mechanism and his notion of "maximizing similarity" in a belief space, using notions like "in the absence of explicit contrary stipulation," and so on. This is very similar to our own statements of the default rule (e.g., Wilks & Bien, 1979), although it does not capture the sort of work we have described here and elsewhere on the strong limitations to the applicability of that rule in conditions of atypical belief (Ballim, 1987; Ballim & Wilks, in press; Wilks & Ballim, 1987). The main point to note is that Fauconnier made great headway with difficult issues such as counterfactuals, presuppositions, and ambiguities of propositional attitude reports by applying an environment-like "mental space" idea.

Of similar relevance is Johnson-Laird's (1983) use of explicit, nested groups of representational items in an application of his mental-model theory of human commonsense reasoning to propositional attitudes. In a different vein, there is a growing amount of work emanating from the modal-logic tradition that is bringing in notions of belief clusters to make the belief logics more accurately reflect commonsense views of belief. See, for example, Fagin and Halpern's (1987) local reasoning logic. It is, however, strange that in this logic it is only in the *semantics* that any notion of clusters is made at all explicit, as "frames of mind." What is important for *reasoning processes* is, of course, clustering made explicit in the representational expressions.

The propositions in John's Reagan environment are not necessarily the ones (about Reagan) that John is aware of, in any sense of "aware" that is closely linked to the ordinary notion of conscious awareness. We are reacting here against the use of the term "awareness" in Fagin and Halpern (1987). The propositions in a belief environment have no necessary relationship to "explicit beliefs" as used by, say, Levesque (1984), because no clear idea is given, by authors using the term, of exactly what explicitness is meant to capture. However, insofar as other authors' explicit-belief notions seem to get at the idea of beliefs agents actually use in reasoning, those notions are exactly our notion of propositions within a belief environment. Our orientation is different, though: We are not interested in massaging modal logic so as to give an appropriate deductive logic of explicit and implicit belief, but rather in devising plausible commonsense-reasoning mechanisms for constructing the explicit-belief sets in the first place.

4. EXTENSIONS TO ViewGen: THE BELIEF ENGINE FIRING ON ALL CYLINDERS

This section reports progress on two extensions to the ViewGen approach: relevance and intensional object identification. Both of these are complex issues that we have not fully resolved, but we can say enough about them to illuminate various other considerations in this article. The relevance subsection gives an idea of the envisaged complexity of relevance determination, and this complexity was appealed to in the earlier section justifying the use of explicit environments. The intensional identification subsection, together with a later section on metaphor, supports the notion that intensional identification, belief ascription, and metaphoric information transfer are three corners of one hat.

It should be noted that, in what follows, we make no firm distinction between beliefs about meaning and beliefs about matters of fact. Hence,

John believes Thalassemia is a province of Greece

reports just another belief (false in this case). Representational consequences follow from this such as that word meanings should also be considered propositional in form, so that they, too, can take part in all the belief-ascription processes we describe. That is no more shocking than noticing that conventional frame representations of meaning can easily be considered to consist of propositions like *Animate(human),* as can any standard form of net representation, linked by set membership and inclusion arcs. And such propositions are clearly about meaning, in some sense, since the fact that humans are animate is hardly a fact about the physical world. As will be seen in Section 7, in treating metaphor we cannot separate issues of fact and meaning.

There would be a considerable philosophical trade-off if we could do away with this conventional distinction: (1) a Quinean one (in the sense of wanting to substitute talk about beliefs and sentences for talk about word meaning; Quine, 1960) where we let the representation of meaning be a function of belief representation, even though this is the inverse of the conventional view; and (2) neo-Quinean, in the sense of aligning ourselves with some current AI-oriented philosophers (e.g., Schiffer, at least in 1972, if not now) who have adopted the view that a self-contained theory of meaning is vacuous, and that such a theory cannot be had independently of a theory of belief and action.

4.1. Relevance

An ascriptional reasoning system must address the issue of relevance simply because, in ascribing a belief or other attitude to an agent, a system should seek to ensure that the belief is relevant to the discourse interpretation needs of the moment. This can involve considerable complexity for a variety of reasons, as will be seen later. Relevance is a complex, variegated notion that has received intense study in its own right, for instance in formal logic (e.g, Anderson & Belnap, 1975), discourse theory (Grosz, 1977; Sperber & Wilson, 1986; Wilks, 1986), AI problem solving (Subramanian & Genesereth, 1987), and elsewhere closer to the present work (Shapiro, 1976; Martins & Shapiro, 1988). Our general strategy, at present, is to seek simple, powerful heuristics that will provide a useful basis for the environment-generation processes that are our current focus.

In the following, we consider the fate of a proposition, *P*, entering the system through the interpretation of natural language input. We assume this proposition is to be taken as a belief of some agent, *A*. We consider the question of whether the proposition should be inserted into a topic environment *E*, for some topic *T*, within *A*'s viewpoint, because of being construed as being relevant to *T*. We assume that, initially, *P* is placed at the top level within *A*'s viewpoint, that is, not inside any particular topic environment. Notice that if *P* is placed inside *E*, it may later be a candidate for pushing into some other environment, and so on.

The overarching strategic question about the role of relevance in our system is about *when* relevance determination is done: To what extent should the determination be "zealous" or "lazy"? A totally zealous approach would consider inserting *P* in *E* as soon as *P* arrives. A totally lazy approach would leave all relevance to be determined on demand; that is, during the course of reasoning about *A's* view of *T*, certain beliefs in *A*'s viewpoint (but outside *E*) would be determined to have become relevant, and therefore to have become candidates for pushing into *E*.

Our approach will be zealous at least to the extent of having a basic rule which zealously deems as relevant those propositions that explicitly mention the topic. Thus, if *T* is John, then the proposition **seriously-ill(wife-of(John))** is relevant. This explicit-mention rule has been the basis of our initial approach to relevance. The presently reported extensions will only account for a limited portion of the full relevance capability that a complete environment-generating system should have. However, they present interesting and significant problems in themselves. A significant problem to be addressed is that of deciding what other manageable and useful types of zeal should be added.

One source for additional zeal is equality statements. Suppose *T* is John, *E* contains a proposition stating that John is Peter's father, and *P* says that Peter's father is seriously ill. Then, surely, *P* is relevant to John and is a candidate for being pushed into *E* zealously, just as much as the proposition stating directly that *John* is seriously ill would be.

Another possible addition of zeal involves inheritance down taxonomic links. Suppose again that John is the topic. Let E state that John is a (medical) patient, and let P say that patients are afraid of the disease thalassemia. Should P be deemed relevant zealously? We suggest that (usually) it should not be, because of the possibly large number of general propositions about patients (and superordinate categories). On the other hand, if the topic were a joint one involving patients in general, as well as John, then P would stand to be deemed relevant anyway, simply by the basic explicit-mention rule. In this specific example we could also consider the possibility of P's being marked as medical expertise, so that it would only be deemed relevant if the agent A in whose viewpoint E lies was believed to be expert on medical matters. Such attention to agent-relative extent of expertise is a feature of the current ViewGen program.

A special case of the taxonomic issue is when, instead of a proposition like the above P—saying patients are afraid of thalassemia—we consider a proposition P that is itself taken to be taxonomic, such as one saying that patients are clients. It may be that such taxonomic information indirectly related to the given topic (John) should be zealously deemed as relevant. The question of how zealously the relevance processing should traverse taxonomic chains is a matter we are investigating.

Inheritance down taxonomies is traditionally concerned with *(quasi-) universal* statements about categories of objects, for example, *all* (or *most*) patients are afraid of thalassemia. However, *existential* statements about categories could also come into play in the relevance issue. Consider a proposition P saying that *some patient or other* in a particular hospital ward is afraid of thalassemia, and suppose John is held to be in that ward. Then P is, in principle, relevant to John (though it need not zealously be deemed to be), because it lends a nontrivial amount of support to the hypothesis that John is afraid of thalassemia. Separate work on *belief convictions* (Ballim, 1988b) will eventually allow investigation of existential statements.

4.2. Intensional Objects and Their Identification

It is natural in a system of partitioned environment notation to treat environments as intensional objects: to treat the Jim-object, pushed down into the Frank-object, as not just yielding by computation an environment that is Frank's-view-of-Jim, but also as a sort of intensional object we might call Jim-for-Frank.[4] Consider now two simple cases of intensional objects to see how the basic default algorithm deals with them:

[4] The names and descriptions attached to environments correspond to the names and descriptions in play in constituent propositions, but we should resist any tendency to think of the environments as being a meaning or referent of the expressions they are named for. The environment names, as far as their meanings go, are simply derivative: dependent, in the best Fregean tradition, on whatever meanings the environment names are assigned on the basis of their participation in the (contained) propositions.

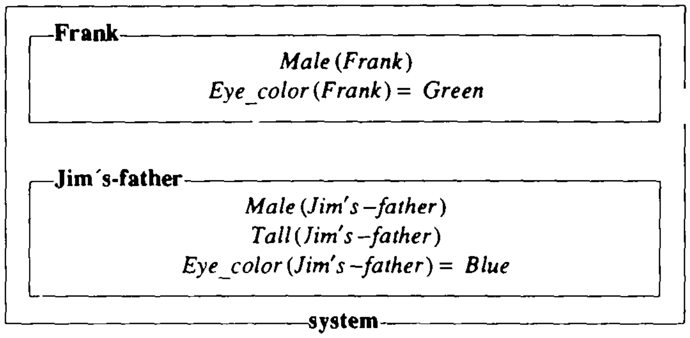

Figure 7. System beliefs about Frank and Jim's-father.

Case 1 (or Two-for-me-one-for-you): The system believes that Frank and Jim's father are two people, but that Mary, whose point of view is being computed, believes them to be the same person.

Case 2 (or One-for-me-two-for-you): Mary believes Frank and Jim's father to be separate people, whereas the system believes them to be the same individual.

Scenarios such as these are common, and arise over such mundane matters as believing or not believing that John's house is the same as the house-on-the-corner-of-X-and-Y-streets.

4.2.1. Two-for-me-one-for-you. Processing of the first case will begin with the system having three topic environments: for Frank, Jim's father, and Mary. Two questions that arise are: What intensional object(s) (i.e., environments) should Mary's viewpoint contain? And what should be the beliefs about those intensional objects? Let us say that the system has beliefs about Frank and Jim's father as shown in Figure 7.

The first question can be rephrased as "given certain intensional objects in one viewpoint (the system, in this case), what are the corresponding intensional objects in the system's version of another viewpoint (Mary's)?" Extending the normal default rule for belief ascription to cope with intensional object ascription, we would say, naturally enough, that intensional objects in one environment directly correspond to identically named (or described) intensional objects in another environment, *unless there is counter evidence to believing this.* This notion of correspondence of intensional objects between environments can be expressed as beliefs, but these beliefs must be of a different type from those previously discussed.

There are two reasons for this: (a) they are beliefs about *intensional (mental) objects*[5] that (b) express the believed relationship between inten-

[5] Beliefs about co-reference are special. Consider the following belief: "John Believes tall(Mike)." This belief expresses that John believes about *the person Mike,* that he is tall. However, the belief "John believes co-ref(Mike, Jim's-father)" expresses a relationship between two intensional descriptions, *not the things of which they are a description.*

sional objects in one space and intensional objects in another space. We represent such beliefs by a predicate called *co-ref*. An occurrence of such a predicate, in an environment about an agent (say, agent *X*), indicates a correspondence between certain objects in the belief space of the agent (say agent *Y*) holding the beliefs about agent *X*, and objects in agent *X*'s belief space. The predicate expresses that the intensional object mentioned for the first person, correspond (as a set) to the intensional objects mentioned for the second person. We are only interested (here) in one-to-one, one-to-many, and many-to-one correspondences. Note that (by default) we assume a one-to-one correspondence. In Section 4.3, we discuss the relationship of co-ref to the more standard "equality" predicate. It should be noted that the correspondence of intensional objects between belief spaces was discussed previously by Fauconnier (1985), Maida (1986, 1988), Wiebe and Rapaport (1986), and Ballim (1987).

In the case at hand (Case 1), Mary's viewpoint ends up containing a single intensional object O (a topic environment) corresponding both to the system's Frank object (topic environment) and to the system's Jim's-father object (topic environment). The question now is to decide what should be put inside the environment *O*. One possibility is to combine the information in the system's Frank and Jim's-father objects *symmetrically,* removing any conflicting information. In the present case, this would result in *O* stating that Frank/Jim's father is male and tall, but stating neither that he has blue eyes nor that he has green eyes. However, we claim that *in realistic situations it will often be more appropriate to take an asymmetrical view,* in which we choose to give precedence either (a) to the information in the system's Frank object over the information in the system's Jim's-father object, or (b) vice versa. Choice (a) reflects the presumption that there is a stronger or closer correspondence between Mary's idea of Frank and the system's idea of Frank than there is between her idea of Frank and the system's idea of Jim's father. This difference of closeness would be plausible, for instance, if the system regarded Mary's view of Frank as being essentially the same as its own except in making the (presumed) mistake of taking Frank to have the property of being Jim's father. Choice (b) reflects the converse presumption, which would be most likely to arise from a hypothesis that Mary is focussing on the person-description "father of Jim," and that she happens to hold that this description identifies Frank. Our claim is that in realistic situations there is more likely to be a reason for making one of these choices than for taking the symmetrical approach.

As an example of such asymmetrical situations arising in discourse, consider the following fragment, in which the boy referred to is Jim.

Mary was listening to what Frank was saying to the boy. It led her to conclude that he was the boy's father.

With reasonable assumptions about the discourse context, it would be apparent that Mary, to some degree, was already knowledgeable about Frank, and was adding to her knowledge the notion that he was the boy's father. This corresponds to asymmetry choice (a) above. To see the potential force of this asymmetry, suppose that the system takes the boy's father to be German, but Frank, American. Then, the asymmetry we are proposing makes the system take the reasonable course of ascribing the "American belief" to Mary, rather than the "German belief." On the other hand, consider the following fragment.

Mary had met Jim's father on several occasions, although he had never told her his name. Under the mistaken impression that he was Frank Timson, she...

It is plausible in this case that Mary was in some respects knowledgeable about the person she thought of as the boy's father, and was augmenting her knowledge with the proposition that this person was Frank. This corresponds to asymmetry choice (b) above. If, again, the system takes Frank to be American, and Jim's father to be German, the asymmetry leads to the reasonable ascription of the "German belief" to Mary.

With either choice of asymmetry, what happens can be affected by the presence of beliefs, that, on the basis of other evidence, the system takes Mary to have had. For instance, if, in the case of asymmetry choice (b), the system has already decided that Mary believed Frank Timson was French, then the imposition of the intensional identification in question should not generally lead to the system going back on its decision. That is, the "French belief" blocks the ascription of both the "American belief" and the "German belief."

The influences on choices of ascription in such examples are more complex than is implied by this brief discussion, but the examples serve to suggest that asymmetry in a particular direction will be well-motivated in many realistic examples.

We handle the asymmetrical choices as follows. For choice (a), the system constructs an intensional object O called "Frank-as-Jim's-father" inside Mary's viewpoint.[6] This object is so-called because it is, so to speak, "the Jim's-father view of Frank" (according to Mary). Notice that this phrase does *not* say that the object is the view of Frank that Jim's father *holds* (according to Mary); rather, the object is a view of Frank that is colored by the idea that he is Jim's father. This way of regarding Mary's intensional object O is directly reflected in the proposed process for constructing O, as will be seen in a moment. Mary's Frank-as-Jim's-father object, O, arises in two stages, as follows (see Figure 8).

Stage 1: The *system's* view of Frank as Jim's father is created. This view is created as a topic environment O' inside the system's viewpoint. The creation occurs in three substages:

[6] There may already be such an object, as we note later.

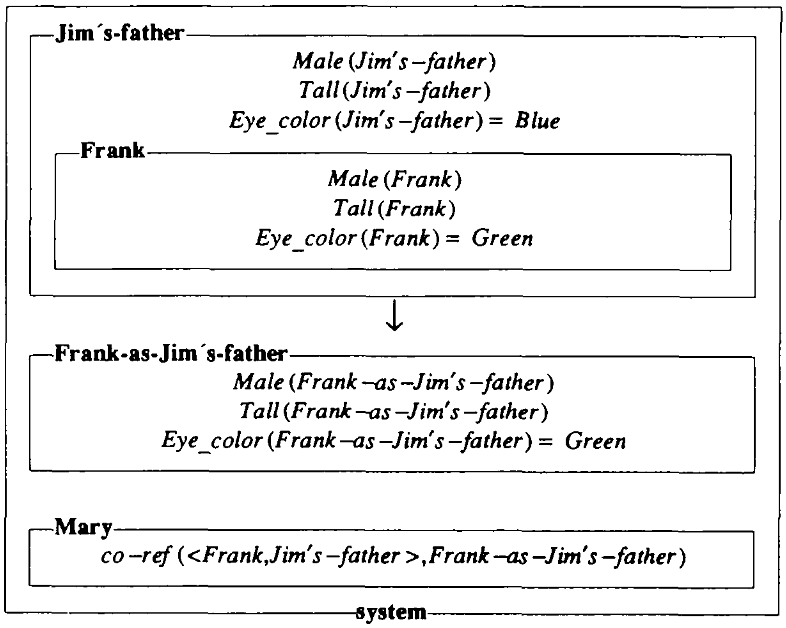

Figure 8. Forming the Frank-as-Jim's-father environment.

1a. Initially, a copy of the system's Frank object (topic environment) is placed inside the Jim's-father object (topic environment). Intuitively, the idea so far is that we have not yet tried to identify Frank as Jim's father, but have merely established a view of Frank that is, so to speak, in the context of Jim's father. That context does not have an effect until Substage 1b.

1b. We now respect the required identification of Frank as Jim's father. We try to push the beliefs in the system's Jim's-father object *inwards* into the Frank object embedded within it, *using the ordinary default rule,* with the slight modification that *Jim's father* is replaced by *Frank* in a pushed belief. Thus, the beliefs that Jim's father is male and is tall are successfully pushed *inwards* (although the former happens to duplicate a belief already in the embedded-Frank object), but the belief that Jim's father has blue eyes is blocked by the green-eye belief already in the embedded-Frank object.

1c. The final substage in constructing the system's Frank-as-Jim's-father object O' is to pull out the Frank object that is embedded within the Jim's-father object, making it into an object (topic environment) O' at top level within the system's viewpoint. In doing this we replace the "Frank" topic-name by the name "Frank-as-Jim's-father," and similarly change the *Frank* symbols inside the environment to *Frank-as-Jim's-father.* The diagram in Figure 8 shows the result, with the arrow notation indicating the pull-out process.

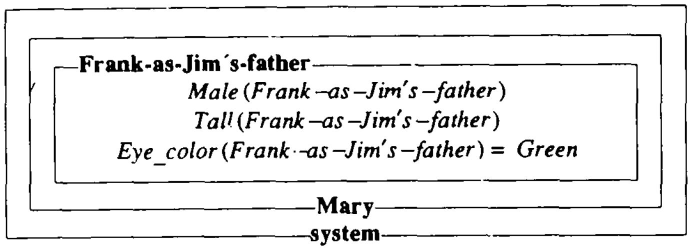

Figure 9. Ascribing the new environment to Mary.

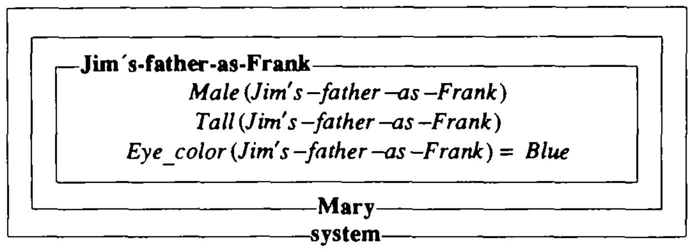

Figure 10. Analogous ascription, with precedence to Jim's father.

Stage 2: We now ascribe the system's beliefs about Frank as Jim's father, that is, the beliefs inside O', to Mary, *once again using the ordinary default rule.* On the assumption that there is no prior information about Mary's view of Frank/Jim's father (e.g., that his eyes are brown), all that will happen is that a copy O of O' will be created inside the Mary viewpoint, giving the revised Mary viewpoint shown in Figure 9. If we had had prior information from discourse input that Mary believes the person's eyes to be brown, then there would already have been a Frank-as-Jim's-father object (topic environment) O inside Mary's viewpoint, and the beliefs in O' would all have been pushed into that object except for the green-eye belief.

If the sytem had decided to give precedence to the Jim's-father information rather than to the Frank information in doing the intensional identification (that is, if it had made choice (b) above) then it would have generated the state shown in Figure 10 by an analogous process.

It might be thought that a symmetric intensional object, with the feature differences appearing as disjunctions (e.g., Eye_color Blue OR Green) would be appropriate as a construct for the Mary environment. We suggest that this is, in fact, psychologically less plausible, and that subjects do construct stronger, and more refutable, hypotheses.

An important thing to notice about the process just described is that the crucial pushing of information from the Jim's-father environment into the embedded-Frank environment (or vice versa) is exactly the type of "inward" pushing used in a particular class of examples with which we illustrated basic belief ascription in Section 2. That was the class where the topic was identical to the believer to whom beliefs were being ascribed. In Sections 5 and 7, we seek to show that belief ascription (e.g., Jim's-father's-view-of-Frank), intensional identification (e.g., Frank-as-Jim's-father), and even metaphor are all different forms of a single fundamental computational process.

The issue of relevance, in the sense discussed in the earlier section, interacts with that of intensional identification in at least two ways. First, if, in the previous example touching upon the ascription of a German or American nationality belief to Mary, it so happened that nationality was irrelevant to the current concerns of the discourse-understanding process, then there would be no need even to address the conflict between nationalities of Frank and Jim's father. This elementary point underscores the importance of devising a good treatment of relevance.

The second point is more complex and remains a matter for further investigation, hinging as it does on the degree of zealousness adopted in dealing with inheritance of potentially relevant information down taxonomic links. We touched upon this type of zealousness in our earlier discussion of relevance. Consider again the choice (b) case of a Frank/Jim's-father situation. Under choice (b), precedence is asymmetrically given to the system's Jim's-father object. Suppose that the system believes that fathers are usually responsible citizens and there is nothing in the system's beliefs about Jim's father that suggests that he is an exception, but on the other hand the system believes that Frank is not a responsible citizen. Assume also that societal attributes are in focus during the discourse understanding. If the system acted zealously with regard to inheritance, it would adopt the explicit belief that Jim's father is a responsible citizen. The system would then ascribe to Mary the belief that Frank/Jim's father is a responsible citizen because of the choice (b) asymmetry.

However, one might argue that intensional identification using specific beliefs, such as that Frank is not a responsible citizen, should be done first, and only then should inheritable defaults be considered. In this example, the belief just mentioned about Frank would be ascribed to Mary, because there would be nothing in the system's beliefs about Jim's father to block it; and, if the system *now* did inheritance, the possible belief that Jim's father is a responsible citizen would no longer be ascribable to Mary.

Under the latter procedure the system could still have proceeded zealously, as long as it had marked its beliefs that Jim's father was a responsible citizen as having been derived by inheritance. It could therefore have been barred from taking part in the specific-belief part of the intensional identification. We suspect that a full treatment of intensional identification will have to pay careful attention to the different types of origin of beliefs.

4.2.2. ***One-for-me-two-for-you.*** In the second case, where the system believes in one individual but Mary two, the natural computation of Mary's

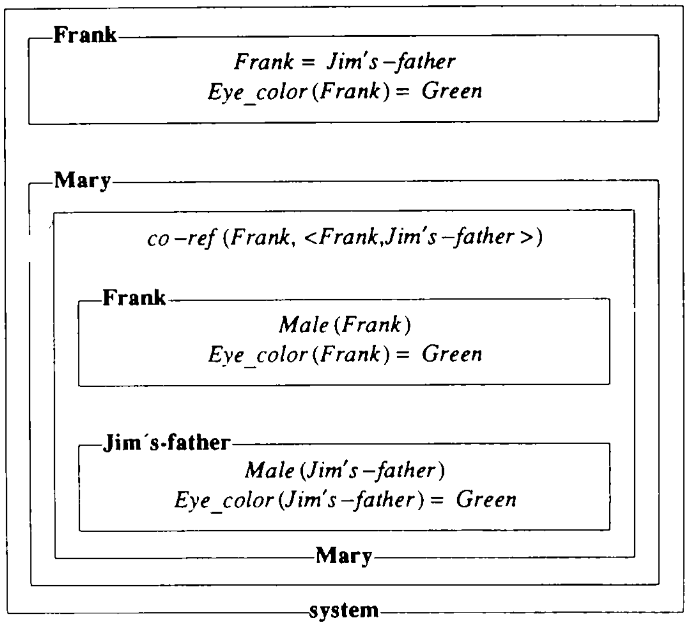

Figure 11. Treatment of CASE 2 on the Frank/Jim's-father example.

view either of Frank or Jim's father is simply to push the system's single representation, changing "Frank" to "Jim's father" as necessary. This is shown in Figure 11.

These are not merely aliases, but are best thought of as dual ascriptions, performed by making two identical copies. Further information about Mary's beliefs would then presumably cause the contents of the two environments to differ, because she presumably has at least some differing beliefs about what she believes to be distinct individuals.

4.2.3. Discussion. Neither Case 1 nor Case 2 turns out to be particularly problematic, and the situation is no different if the entities about whose identity there is dispute are nonbelievers rather than believers. Those would be like the classic but natural cases such as a difference between dialogue participants as to whether Tegucigalpa and Capital-of-Honduras are, or are not, the same, or as to whether Rome or Avignon should be identified with City-of-the-Popes.

More difficult cases, which bring in all the panoply of philosophical distinction and discussion, are those conventionally discussed under the *de re/de dicto* distinction. One type is the following: The system reasonably believes Feynman to be a famous physicist but encounters Frank who, on the strength of a single appearance on the TV screen, believes Feynman to be a famous TV performer. For the sake of this example, it is essential to accept that the two occupations are incompatible. Suppose the discussion now forces the system to construct its view of Frank's view of Feynman. Now, there will be no point at all in performing that computation unless the system believes Frank's beliefs to be *de re.* Frank no doubt considers his own beliefs *de re,* as we all do. The crucial thing is that the system believe this, and the test would be some proposition in the Frank environment, and ABOUT Frank, equivalent to *("Feynman" names Feynman).* If that is not present, the system should infer that Frank has another person in mind: that his beliefs are *de dicto* FOR THE SYSTEM, and hence any push-down computation would be pointless.

Consider the relation of this example to the former, simpler cases, where the system can identify or separate distinct environments. This last case would be similar if the system knew which non-Feynman individual Frank was confusing Feynman with, perhaps Johnny Carson. In that case, the system could perform a push down, even though it believed Frank's beliefs to be *de dicto* as far as Feynman was concerned, for they would be *de re* with respect to Johnny Carson. The system could then push Carson into Frank, while changing the resulting environment's name to "Feynman." To summarize, the absence of *("Feynman" names Feynman)* in the Frank environment is only a reason for not pushing down Feynman, but leaves open the possibility of some other *de re* push down.

4.3. Co-reference versus Equality

A special point about intensional identification (and relevance) arises from the issue of equality versus co-reference, where the former is the deeming of referents as identical and the latter the deeming of (different) intensional descriptions as co-referential. Our use of environments corresponds naturally to the use of intensional entities deemed co-referential, and hence, to the implicit use of a co-reference (rather than equality) operator. In that sense our assumptions are very like those of the CASSIE group (Maida & Shapiro, 1982; Shapiro & Rapaport, 1986) except that we see no need to make any strong claim, as they do, that only co-reference will ever be used, and that all entities in the system are intensional. The crucial point in our system is that the environment notation moves, as it were, the belief predicate, at any level of nesting, out to the environment boundary or partition, and so, within an environment, we have *precisely the conditions of a belief space that sanction substitution of co-referents without problems,* as in the *de dicto/re* examples above.

The use of co-reference statements linking terms denoting intensions, as in **co-ref(Father-Of(Peter), Boss-Of(Jim))**, has a well-known advantage over the use of equality statements linking the corresponding ordinary terms, for example, **father-of(peter) = boss-of(jim).**[7] The advantage is that the co-reference statements allow more controlled separation of inference about a thing under different descriptions than the equality statements do; and the separation gives us, in turn, an extra, *explicit* handle on relevance (Section 4.1). Since co-reference statements do not sanction substitution in the way that equality statements do, we could have the expression **Strict-Boss(Boss-Of(Jim))** without being automatically tempted to produce the expression **Strict-Boss(Father-Of(Peter)). (Strict-Boss** is a function that takes a person-concept and delivers a concept of that person being a strict boss.)

We could view all this as having special axioms that sanction co-reference-based substitutions only under certain conditions, rather than having to adopt a nonstandard meaning for the equality predicate or having knowledge-intensive, behind-the-scenes heuristics that limit the application of equality-based reasoning. For instance, we could have an axiom schema of the (very rough) form:

> ***P*(*T*) and co-ref(*T*, *U*) and *C*→*P*(*U*)**
> provided that: *P* is an "intensional predicate" in a domain *D1*, *T* is an intensional term describing something using the resources of *D1*, *U* is an intensional term describing something using the resources of some domain *D2*, and *C* is a formula stating that the system is currently considering cross-inferences between *D1* and *D2*.

If *D1* and *D2* are the employment and family domains respectively, then an example of *P*, *T* and *U* could be **Strict-Boss, Boss-Of(Jim),** and **Father-Of(Peter).** What we would need behind the scenes is a single heuristic giving lower priority to equality-based reasoning than to co-reference-based reasoning.

However, there is no need for such an axiom schema if we know our inferences are limited to the appropriate environments; that is precisely what our partitioning provides, as, in principle, do all systems that look back to Hendrix's (1979) partitioned networks, although his work, of course, provided no analogue of belief ascription.

5. METAPHOR: SHIFTING THE BELIEF ENGINE TO A HIGHER GEAR

Metaphor is normally explicated, formally or computationally, by some process that transfers properties by some structural mapping from one structure (the vehicle) to another (the tenor). A classic example in AI would

[7] Here we are appealing to the notation of Creary (1979), where the noncapitalized symbols denote ordinary subjects, functions and predicates in the domain, whereas the capitalized symbols denote intensional objects and functions. For instance, **Boss-Of** is a function that takes a person-concept and delivers a concept of that person's boss as such.

be the work of Falkenhainer, Forbus and Gentner (1986), and Indurkhya (1987). All these authors were concerned, as we are, with metaphor and analogy viewed as some form of structural mapping; the difference between what they offered and what we offer here is the linkage between that process and those of intensional identification and belief ascription. Some would object here about the issue of transferring properties versus transferring structure, but we shall not enter this argument here because, although the following examples transfer properties within propositional beliefs, it will be clear from the discussion in Section 7 that we consider our current representation inadequate and only illustrative, and that a fuller representation would display mapping of more complex structures. Again, in this section we shall play fast and loose with the metaphor versus metonymy and the metaphor versus analogy distinctions. For our purposes here, those distinctions affect nothing.

We are exploring the possibility of applying our basic belief algorithm to metaphor, as an experiment, to see if it gives insight into the phenomenon. That should not be as surprising as it may sound: Metaphor has often been viewed, in traditional approaches, as "seeing one thing as something else," a matter of viewpoints, just as we are presenting belief. We propose that propositions in the topic environment for the vehicle of a metaphor be "pushed inward" (using the standard algorithm, presented before), into an embedded environment for the tenor, to get the tenor seen through the vehicle, or the view of the tenor-as-vehicle. This process was already described in Section 4.2 on intensional identity.

The key features here are:

1. One of the conceptual domains is viewed as a "pseudobeliever".
2. The pseudobeliever has a metaphorical view of a topic or domain.
3. The generation of such a view is not dissimilar from ascribing beliefs by real believers.
4. Explicating this by pushing or amalgamating environments yields new intensional entities after an actual transfer of properties.

So, in the classic historical case of atom-as-billiard-ball, given the environments for atom and billiard ball as shown in Figure 12, we generate the environment for atoms as billiard balls as follows. The environment for atoms is nested within the environment for billiard balls, and then the contents of the billiard ball environment are pushed down into the nested-atom environment, replacing the term "billiard ball" by "atom" wherever it occurs in propositions being pushed. The overriding of properties would follow in the same way as for standard beliefs: For example, **Light(atom)** overrides the incoming **Heavy(billiard ball)**. However, **Round(billiard ball)** would survive as the property **Round(atom)**—correctly for the original analogy—because there would be no preexisting shape property in the system's belief set for atoms. Then, the nested-atom environment is pulled out to form a

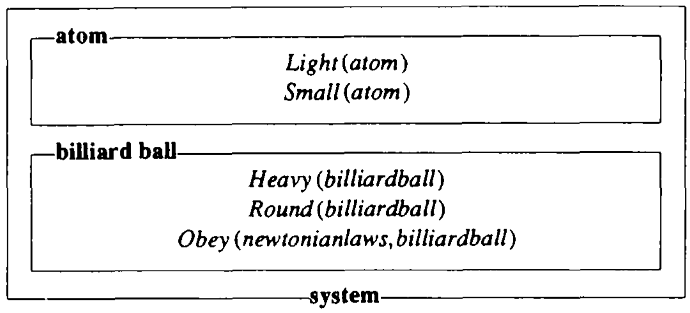

Figure 12. System beliefs about atoms and billiard balls.

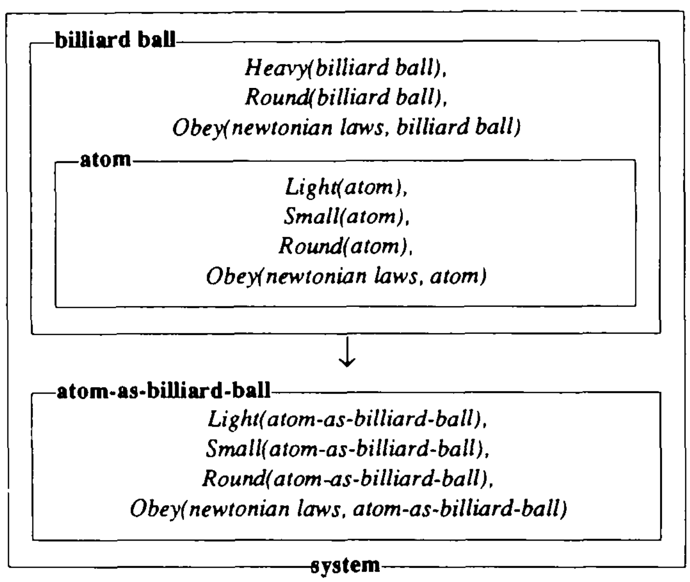

Figure 13. Forming the atom-as-billiard-ball environment.

new environment "atom-as-billiard-ball," replacing such occurrence of "atom" with "atom-as-billiard-ball." This new environment is the metaphoric view of atoms as billiard balls. Figure 13 uses an arrow, as before, to illustrate the process. Similarly, in

Jones threatened Smith's theory by reimplementing his experiments.

we would know we had a preference-breaking, and potentially metaphorical, situation from the object-feature failure on "threaten" (which expects a person object). Or, rather, Wilks (1977) argued that metaphors could be identified, procedurally at least, with the class of preference-breaking utterances (where, in a wide sense, assertions relating two generic classes, as in "An atom is a billiard ball" or "Man is an animal," can be preference breaking). The awkward cases for that broad delimitation are forms like "Connors

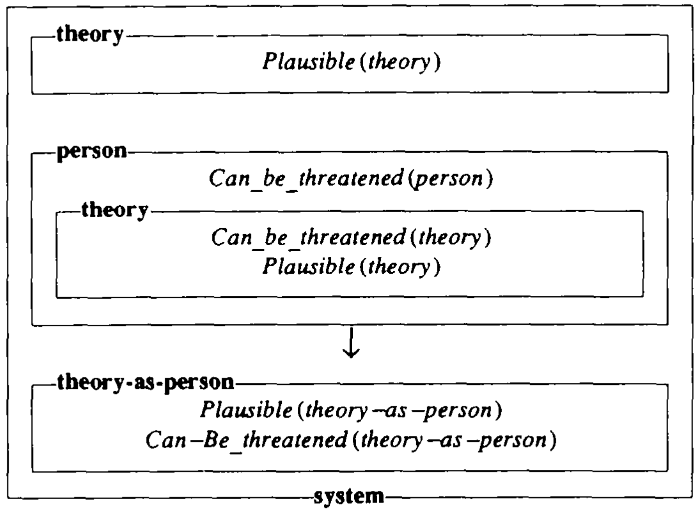

Figure 14. Forming a theory-as-person environment.

killed McEnroe," which breaks no verb preferences but is read metaphorically by some as "beat soundly at tennis." Here, one might consider taking the classic Marcus escape and using our procedural definition to rule this example out of court as a "garden-path metaphor." However, as we shall see in later discussion (Section 7), there is a deeper way in which preferences and metaphor are linked.

We could now plausibly form a metaphoric view of theory-as-person using the same process as above, and using the assumption that the basic preferences of the concept "threaten" (Wilks, 1977) are for a person as agent and as object (if that is not accepted, a metaphorical push down can begin from whatever such preferences an objector would be prepared to assign to the action.) Figure 14 shows possible system environments for theory and person, and the resulting theory-as-person environment.

By this maneuver, a new and complex metaphorical property of theories is derived. It might be, of course, that this procedure of belief overriding as a basis for metaphor would produce no different a set of plausible properties transferred than any other system (e.g., Falkenhainer et al., 1986); that would be, again, an experimental question. But its importance or originality would lie in the fact that it was further application of an algorithm designed to explicate another phenomenon altogether (i.e., belief) and, therefore, yield a procedural connection between the notions, one that has other intellectual justifications, as will be shown in a moment.

In principle, the method should extend to other phenomena widely considered to be metaphorical (Cohen & Margalit, 1972) but with a quite different grammatical basis, such as "rubber duck." Here, we can envisage the push down of environments (duck and rubber object), after which properties like animacy from the DUCK environment would be canceled by the preexisting property (alias belief) "inanimate" within the RUBBER environment so that we did not end up with rubber ducks (alias rubber-as-a-duck) being animate. Cohen and Margalit argued that there could be no principled basis for property transfer in metaphor explication, but, in a sense, all computational accounts, including this one, consider this an empirical claim, one which AI researchers believe is false. Here, the principled basis would fall back on a relevance algorithm (see Section 4.1) supplemented by the default-belief algorithm. The intuitive support for what we propose comes from a deep connection between belief and metaphor: taking metaphor-as-false-belief (Davidson, 1978) seriously, in that metaphors for a particular believer are just special beliefs, ones which can, of course, become more generally believed (e.g., Men are beasts! Women are cats!).

There is a further interesting aspect to the connection between belief and metaphor. We have stressed a procedural connection that may seem improbable to some people. There is also the important but neglected phenomenon of the content of belief being inherently metaphorical, and in a way that conventional theorists totally neglect by their concentration on simplistic belief examples like "John loves Mary." A far more plausible candidate might be a truth such as:

Prussia threatened France before invading it successfully in 1871.

What are we to say of this historically correct belief? What are the entities referred to by "Prussia" and "France"? Simple translation into some first-order expression like **Invade(Prussia, France, 1870)** just obscures the real problem, one for which the semantics of first-order logic are of no help at all. Are the entities referred to somehow metaphorically the Prussian people, or army, or a part of the army?

Following the approach described earlier, we might expect to detect breaking of linguistic preferences of the verb "threaten" and perform a trial push down of properties of the "people" environment (given by the conventional preferences of "threaten") into an environment for Prussia (= a land mass, the basic representation). An important safeguard, which there is no space to discuss here, would be that we examine our inventory of representations to see if we have one for "Prussia" that already expressed the (dead) metaphor of a country-name-as-a-polity (some would insist that this was a metonymy, but we decided not to make this a significant issue in this article).

The possibility of a metaphorical belief belonging to some agent other than the system itself underscores the benefits of our method of unifying metaphorical transference with belief ascription. Suppose that (according to the system) Bill has a certain metaphorical belief *B*, perhaps the Prussia/France one if we assume that it does indeed involve a country-as-person metaphor. Suppose, now, that the system takes Bill to attribute this belief *B* to Sally in the ordinary default way (subject to the check that, according to Bill, Sally is "qualified" to have beliefs about European history). Here we have a combination of metaphorical transference (from the person domain to the country domain) and belief ascription: A combination that could, of course, appear at any level of nesting of belief. These processes essentially work by the same mechanisms in our method. This obviates the need that would otherwise exist to create a suitable interface between mechanisms for belief ascription and (previously unrelated) mechanisms for metaphorical transference.

Furthermore, we may anticipate a point we make later about fuzziness of the distinction between intensional identification and metaphor. Our method allows the system to be neutral as to whether a belief *B* of an agent Bill is viewed *by him* as being metaphorical or not. If it is, then the environment manipulation involved in constructing, say, a country-as-person environment is consistent with the system taking Bill to be thinking metaphorically. On the other hand, if Bill does not regard *B* as metaphorical, then the same environment manipulation is consistent with the system taking Bill to be (partially) confusing the notion of a country with the notion of a person, and thus performing an intensional identification (albeit one between general concepts rather that concepts of individuals). With our method, there is simply no need for the system to have to adjudicate on whether Bill is engaged in metaphorical thinking or not.

Notice, finally, that in the environment-based processing of metaphor there is an asymmetry of available push down, much as with the construction of intensional entities in an earlier section. This asymmetric duality of metaphor is exactly that of the alternative treatments of:

My car drinks gasoline.

in Wilks (1977), where one can consider the statement as being a car-as-drinker metaphor OR as a drinking-as-using metaphor, and only overall coherence with a database of cases and knowledge structures will tell one which is right. In that work, the model was not one of beliefs, but (in the spirit of its age) of framelike structures expressing dictionary information. But the underlying point is the same: Preference violations are the cues or triggers for metaphorical processes but do not settle which metaphor (depending on the directionality of the preferences) should establish itself in the context. That is a matter for other, more general inference processes and the coherence of what they produce. In the examples here, we have simplified the matter by considering only a single push down for each example.

6. TOWARDS A GENERAL THEORY OF SPEECH ACTS

Much work has been done in recent years in developing natural language processing (NLP) systems that interpret sentences in terms of speech acts[8]

[8] See Austin (1962) and Searle (1969).

(Allen & Perrault, 1978; Cohen & Levesque, 1987; Perrault, 1987). As noted earlier, the relation of our basic belief-ascription method to that work is that those authors assumed some partition of the beliefs needed for understanding into viewpoints and to any required depth of nesting. That is to say, they assumed those environments were already in existence as a database before speech-act computations were done. In our view, this is psychologically and computationally unrealistic and, for us, the creation and maintenance of nested viewpoints is the primary computational and theoretical task. Seen in that way, we are not so much building on their work as providing a foundation for it, by building a processing model of their key assumption.

Our approach can thus be seen as (a) a demand for more realistic complexity in belief-environment computation and, at the same time, (b) a reaction against the complexities of speech-act analysis in, for example, the work at the University of Toronto (and we believe that Perrault, 1990, made this latter move, too). If we treat belief less simplistically, we get a simpler treatment of speech acts as a reward. But our main assumption in treating speech acts is similar to that of the other approaches mentioned: We locate a belief environment, usually of the beliefs of the system about the beliefs of another agent about the system itself, within which reasoning is done so as to make sense of otherwise incomprehensible dialogue input. This most general assumption also serves to link the treatment of speech acts to that of metaphor: A belief environment is created that "makes sense" of otherwise anomalous input. However, by including speech acts in this article, we intend only this connection of ideas, and not that speech acts are phenomena that, like metaphor and intensional identification, can be seen as modeled by the same process as belief ascription.

As many commentators have pointed out, the construction of plans corresponding to speech acts on each occasion they are encountered is implausible. For example, it would be inefficient to work out that the surface interrogative:

Can you give me your departure time?

was a request *each time it was encountered.* In our view such "speech-act interpretation shifts," which do not undergo significant changes over time in a language, are best seen as stored, learned wholes.

All this is purely programmatic, but we are concerned here with establishing that speech acts are part of a family of notions, along with intensional identification, metaphor, and belief itself that are inseparably linked. It is not only that speech acts rest upon some belief calculus for their formal expression, but that speech-act phenomena themselves are not always separable from metaphor, say.

A real and interesting murder case in Britain in the 1950s concerned robbers called Craig and Bentley. Craig shot and killed a policeman on a roof, but was 16 and not hanged. Bentley was hanged for shouting at Craig

"Let him have it." His (unsuccessful) defense at this trial was that he intended by those words that Craig should give the policeman the gun but was misunderstood. Guilt being (in theory) based on intention rather than causality, that was a reasonable defense, whether or not it was honest. It was part of his defense that he intended the literal meaning of the words and not the (conventionalized or dead) metaphor "shoot." Clearly, both alternatives admit of similar speech-act analysis, but the interesting issue relevant here is under what conditions the beliefs in an environment lean towards an interpretation of input as metaphorical (by some such methods as we have discussed) because that would be a determination prior to any determination of what speech act was in play.

7. THE GENERAL ISSUE OF BELIEF, INTENSIONAL IDENTIFICATION, AND METAPHOR

The goal of this article has been the application of notions derived for belief to the explication and modeling of intensional entities and metaphor understanding. In this section we recap our views both on this idea and on the other fundamental links between belief processing and metaphor. First, we summarize our views on the question of how, in our view, intensional identification fits with both belief ascription and metaphor.

7.1. Belief Ascription and Intensional Identification

Intensional identification intrinsically involves some sort of combination of two or more bodies of information, whether or not one follows our environment-based approach. We also claim that intensional identification is likely to have an asymmetrical quality as a matter of fact, and this makes the asymmetric aspect of belief ascription a plausible technique for constructing the intensional entities. In the Mary example, this might be because Mary's Frank/Jim's-father idea is likely to correspond more closely to one of our two person-ideas than to the other, and we might also expect there to be dialogue clues from which we could infer Mary's presumed direction of conflation. This is not to deny the possibility of more complex situations where there is no clear precedence, but the approach is a heuristically plausible one.

7.2. Intensional Identification and Metaphor

The identification of intensional objects *A* and *B* (done with bias towards *A*) is a matter of taking *A as B*. We hold that this "as" is the same as in taking a metaphorical target *A as* the vehicle *B* of the metaphor (e.g., atom as billiard ball). In both cases, one view is imposed upon another (information about *B* is imposed upon *A*). This correspondence does not amount to saying that there are no differences between typical intensional identification and typical "metaphorizing." Certainly, the latter is likely to involve more unusual, unexpected, or category-crossing impositions of information.

Nevertheless, the two processes are similar, both conceptually and from the procedural point of view of the detailed computational processes taking place. Moreover, in cases where someone uses a phrase like "God the Father," we might not be able to say whether that was an example of the conflation of two (independent) intensional entities, or a metaphor. The method of this article suggests that, if the basic computational technique were the same for treating both, we would not have to decide that question.

7.3. Belief and Metaphor

Here we return to the core idea of this article, namely, that representational and processing notions derived for belief can usefully be applied to the explication and modeling of metaphor understanding. The core idea has a general force derived from the fact that metaphor has often, in the literature, been seen as a point-of-view phenomenon, or "seeing something as something else." But all that is very general support: The crucial idea here has been the application of a precise notion of computational belief ascription to metaphor, and transferring properties (expressed as believed propositions) by our standard algorithm in order to create a metaphorical point of view of an entity.

One type of analogy that can be drawn in mundane discourse is between different people's states of mind or belief frameworks. Consider, for instance, the discourse fragment "Bill is a chauvinist...John is like Bill." Assuming there is no interruption of coherence here, the reported analogy between John and Bill is one of belief framework. That is, chauvinist beliefs of Bill's can be transferred (by default) to John. What we have here is straightforward belief ascription that is *also* a case of analogical tranference, which is essentially the same thing as metaphorical transference. This intersection provides considerable additional support to the basing of metaphorical transference on the extended belief-ascription mechanism.

However, we also wish to mention, although there is no space here to defend it fully, the force and generality of the converse notion: *Belief ascription, as a fundamental psychological and computational process, is also logically and empirically dependent on metaphor.*

In one sense, that claim is trivial, because all computational approaches to propositional attitudes ultimately rest on underlying metaphors: Most commonly, metaphors that bring in the idea of "possible worlds" or "situations," or others that cast the mind as holding, possessing, or being otherwise related to abstract objects akin to natural language sentences or logical formulae. Our approach rests on a metaphor in the latter class, namely the mind-as-container metaphor, under which the minds and belief sets of others are seen as porous containers that can be nested like buckets or jars. This metaphor carries with it the explicit grouping idea emphasized in Section 3.

But we intend something much more general here, independent of any particular prevalent metaphor for the mind or belief states. First, consider the precept that, in plausibly hypothesizing what some agent *X* believes on some topic *T*, one proceeds largely by trying to ascribe one's own beliefs about *T* to *X*, perhaps failing to do so because of contrary beliefs about *T* one already knows *X* to have. What we are now suggesting is that this activity is very much like metaphorizing: the process of "ascribing" information from the metaphor vehicle to the tenor, perhaps failing to do so because of contrary existing tenor information that one wishes to preserve. Specifically, in a belief-ascription activity *one uses one's current belief state about the topic T as the vehicle of a metaphor, the target being the other agent's belief state.* In brief: One uses one's own state of mind as a metaphor for other people's. This has a general similarity to Maida's (1986) view.

A second very general aspect of the dependence of belief processing on metaphor can be seen by considering the unexamined assumption we have made throughout this article, which is also one that virtually all AI researchers and logicians use for discussing beliefs: Beliefs can be conveniently expressed as simple propositions, which contain predicates, which unfortunately look like words, but, in fact (so the assumption goes), univocally denote entities that are concepts or world referents.

Everyone knows that this assumption, underlying all modern formal semantics as it does, is a claim of highly dubious content, and it is particularly so if we consider the fact—always cited in the research of Wilks (e.g., 1977) on preference semantics—that many, if not most English sentences in real texts like newspapers, are preference breaking: That is to say, the concepts contained in them are used out of their dictionary-declared contexts of constraint, as in "Prussia attacked France." This is no more than a repetition of the now common observation that much normal discourse is "metaphorical" in a broad sense, but what is not so often concluded, as it must be, is that this has strong and destabilizing consequences for any formal semantic representation of language (cf. Johnson, 1987; Lakoff, 1987), and for belief ascription in particular.

In the face of such observations, the notion of univocal predicates as the basis of formal representations of a natural language, freed from the contamination of languages like English, becomes hard to sustain, and the problem is in no way solved by allowing for non-univocality (i.e., indexing predicates for particular dictionary word senses; e.g., POST1 and meaning only a stick) because the ubiquity of metaphor or preference-breaking use suggests that a natural language is used normally and comprehensibly even when no such indexing to conventional senses can be done. And, it should not need adding, this difficulty is not alleviated at all by those who say things like "we do not use predicates, only axiomatic structures, or sets of n-tuples." To them, the answer is simply that the only way they have of knowing which set or axiom is which must be by means of the associated predicate name, and then the above problems return unsolved.

If we now return to our central theme and consider that those comprehensible sentences, containing non-sense-indexible metaphorical uses, are the stuff of beliefs also, and that they must also be ascribed by believer to believer, then what trust can we put in the sorts of naive representations used in this and every other article on the subject? The short answer is none, unless we can at least begin to see how to move to a notion of representation of meaning for belief ascription that also takes the metaphoricity of beliefs and language as basic.

At present, we can do little more than draw attention to this phenomenon, so that we cannot, in the future, be accused by our successors of more naivety than necessary. However, we believe we know where to look, and what other aspects of current research to draw into work on belief ascription. One essential for the future is to link the present work fundamentally to work on meaning that is both dictionary based and shows how to extend beyond that, so that new usages can be represented, usually within networks of associations as the basis of discrete senses (Fass, 1987; Wilks et al., 1989, in press). Another essential is that the sorts of explorations we have carried out here on explicating the notion of metaphor via belief ascription be bootstrapped back into the belief-ascription process, so that we can ascribe a belief from believer *A* to believer *B* that "Smith attacked Jones's notion of continuity" in such a way as to assume that the metaphorical content of "attack" here also transfers from environment to environment (saving here the assumption that culturally similar believers may be assumed to have the same metaphorical processing mechanism, just as they do the same belief-ascription mechanism. But those assumptions, too, might have to be relaxed in certain situations). Such transfers are central to work by Barnden (1989a, 1989b, 1990).

One interesting class of cases of this phenomenon will be those where a system believes that another believer has false (as opposed to metaphorical) beliefs about word meaning. To return to the believer who thinks Thalassemia is a province of Greece, he is confronted by the input phrase "The cure for Thalassemia." A system might predict that, faced with what should be a radical preference violation, the believer will give up and ask for help, and so the system might wait and see and make no ascriptions. But a plausible zealous strategy would be to ascribe the results of a metaphorical push down (based, in the system's own view, on wholly false beliefs about meaning). Anyone who considers this implausible should consider the locution, heard recently on American television, "The cure for Panama."

If we can escape from the basic representational assumption, made here and everywhere else (because it is so hard to think of anything else!) that the predicates in the (ascribed) representation for belief are sense-determinate in some simple denotational way, then the problem may be soluble, and require, as we noted, only some method of metaphor processing (by belief-like methods such as those we propose here) during the belief-ascription process.

An alternative, and lazier, possibility is that we move to a representational phase where we make no strong referential assumptions about the meanings of the predicates in beliefs ascribed from believer to believer (just as one can assume that if natural languages are very close, like Dutch and German, we may not need to sense-resolve words transferred between them, allowing the target understander to do the work). Then we could use a process like the metaphor processor described here only on demand, when required to push an interpretation below/beyond its metaphorical expression. This again is consistent with certain strong and plausible assumptions about human processing. Whichever of these alternatives is ultimately chosen, both require recognition of the intimate dependence of belief ascription on the metaphoricity of language and belief representations.

8. CONCLUSION

This article advocates a highly "pragmatic" approach to propositional attitudes. Rather than being concerned with traditional issues such as devising an elegant axiom set, satisfying semantics, or adequate proof procedure for a belief logic, we believe that concern should be focused on a commonsense plausible reasoning schema about propositional attitudes. In particular, we are interested in ascriptional reasoning about attitudes. We claim that for ascriptional reasoning, it is important to concentrate on environments: groups of propositions that can be manipulated as explicit units, rather than as implicit groups arising only behind the scenes.

Our main concern has been to demonstrate some of our reasons for thinking that belief processing and metaphorizing are strongly interdependent, and indeed very similar in some respects. The essence of metaphorizing is assimilable into a generalization of the environment-manipulation procedures we originally devised for handling ordinary belief ascription. Conversely, belief ascription is, in large measure, assimilable into metaphorizing, in that one's ascriptional activities use one's states of mind as metaphors for other people's states of mind. Moreover, Barnden (1989a, 1990) argued that metaphors for the mind, which are commonly used by people in ordinary discourse, have to be given a central role in representational approaches to propositional attitudes. Our ViewGen work already observes this to a useful extent by adopting, via environments, the prevalent mind-as-container metaphor.

We have also presented our reasons for perceiving deep connections between intensional identification on the one hand, and both belief ascription and metaphorizing on the other. Part of our view is a claim about intensional identification being typically asymmetrical. A corollary of these connections is a strengthening of the bond between belief processing and metaphor. We are investigating the extension of our approach to deal with speech acts, and the incorporation of a sophisticated but heuristically restricted treatment of relevance. The expense of relevance processing is one reason for wanting to use explicit groupings of beliefs. Finally, we resist the possible objection that our linkage of belief to metaphor requires the problem of metaphor to be fully solved first: a huge task. Rather, research on metaphor to date can serve as a basis for useful progress with belief processing, and vice versa.

REFERENCES

Allen, J.F., & Perrault, C.R. (1978). Participating in dialogue understanding via plan deduction. *Proceedings of the Second National Conference of the Canadian Society for Computational Studies of Intelligence*. Toronto, Canada.

Anderson, A., & Belnap, J. (1975). *Entailment: The logic of relevance and necessity*. Princeton, NJ: Princeton University Press.

Austin, J.L. (1962). *How to do things with words*. Oxford, England: Oxford University Press.

Ballim, A. (1986). Generating points of view. In *Memoranda in computer and cognitive science* (Tech. Rep. No. MCCS-86-68). Las Cruces: New Mexico State University, Computing Research Laboratory.

Ballim, A. (1987). The subjective ascription of belief to agents. In J. Hallam & C. Mellish (Eds.), *Advances in artificial intelligence*. Chichester, England: Wiley & Sons..

Ballim, A. (1988a, July). A language for representing and reasoning with nested belief. *Proceedings of the First Annual Irish Conference on Artificial Intelligence and Cognitive Science*. Dublin, Ireland.

Ballim, A. (1988b). *Belief convictions, quantification, and default reasoning*. Unpublished manuscript.

Ballim A., & Wilks, Y. (in press). *Artificial believers*. Hillsdale, NJ: Erlbaum.

Barnden, J.A. (1983). Intensions as such: An outline. *Proceedings of the 8th International Joint Conference on Artificial Intelligence* (pp. 347-353). Los Altos, CA: Morgan Kaufman.

Barnden, J.A. (1986). Imputations and explications: Representational problems in treatments of propositional attitudes. *Cognitive Science, 10*, 319-364.

Barnden, J.A. (1987a). Interpreting propositional attitude reports: Towards greater freedom and control. In B. du Boulay, D. Hogg, & L. Steels (Eds.), *Advances in artificial intelligence-II*. Amsterdam: Elsevier (North-Holland).

Barnden, J.A. (1987b). Avoiding some unwarranted entailments between nested attitude reports. In *Memoranda in computer and cognitive science* (Tech. Rep. No. MCCS-87-113). Las Cruces: New Mexico State University, Computing Research Laboratory.

Barnden, J.A. (1988). Propositional attitudes, commonsense reasoning, and metaphor. *Proceedings of the 10th Annual Conference of the Cognitive Science Society* (pp. 347-353). Hillsdale, NJ: Erlbaum.

Barnden, J.A. (1989a). Towards a paradigm shift in belief representation methodology. *Journal of Experimental and Theoretical A.tificial Intelligence, 1*, 131-161.

Barnden, J.A. (1989b). Belief, metaphorically speaking. *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning* (pp. 21-32). San Mateo, CA: Morgan Kaufmann.

Barnden, J.A.(1990). Naive metaphysics: A metaphor-based approach to propositional attitude representation. In *Memoranda in computer and cognitive science* (Tech. Rep. No. MCCS-90-174). Las Cruces: New Mexico State University, Computing Research Laboratory.

Carbonell, J.G. (1982). Metaphor: An inescapable phenomenon in natural language comprehension. In W. Lehnert & M. Ringle (Eds.), *Strategies for natural language processing.* Hillsdale, NJ: Erlbaum.

Charniak, E. (1978). On the use of knowledge in language comprehension. *Artificial Intelligence, 11,* 225-265.

Cohen, P., & Levesque, H. (1980). Speech acts and recognition of shared plans. *Proceedings of the Third Biennial Conference of the Canadian Society for Computational Studies in Intelligence* (pp. 263-271).

Cohen, P., & Levesque, H. (1985). Speech acts and rationality. *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics* (pp. 49-60). University of Chicago.

Cohen, P.R., & Levesque, H.J. (1987, July). Intention = Choice + Commitment. *Proceedings of the Sixth National Conference on Artificial Intelligence,* 410-415. Seattle.

Cohen, J., & Margalit, A. (1972). The role of inductive reasoning in the interpretation of metaphor. In D. Davidson & G. Harman (Eds.), *Semantics of natural language.* Dordrecht, Holland: Reidel.

Creary, L.G. (1979). Propositional attitudes: Fregean representation and simulative reasoning. *Proceedings of the Sixth International Joint Conference on Artificial Intelligence* (pp. 176-181). Los Altos, CA: Morgan Kaufmann.

Davidson, D. (1978). What metaphors mean. *Critical Inquiry, 5,* 31-48.

Dinsmore, J. (1987). Mental spaces from a functional perspective. *Cognitive Science, 11,* 1-21.

Fagin, R., & Halpern, J.Y. (1987). Belief, awareness and limited reasoning. *Artificial Intelligence, 34,* 39-76.

Falkenhainer, B., Forbus, K., & Gentner, D. (1986). The structure-mapping engine. *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 272-278). Los Altos, CA: Morgan Kaufmann.

Fass, D.C. (1987). Collative semantics: An overview of the current Meta 5 program. In *Memoranda in computer and cognitive science* (Tech. Rep. No. MCCS-87-112). Las Cruces: New Mexico State University, Computing Research Laboratory.

Fauconnier, G. (1985). *Mental spaces: Aspects of meaning construction in natural language.* Cambridge, MA: MIT Press.

Grosz, B.J. (1977). The representation and use of focus in a system for understanding dialogs. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence* (pp. 67-76). Los Altos, CA: Morgan Kaufmann.

Hendrix, G.G. (1979). Encoding knowledge in partitioned networks. In N.V. Findler (Ed.), *Associative networks.* New York: Academic.

Hobbs, J.R. (1983a). Metaphor interpretation as selective inferencing: Cognitive processes in understanding metaphor (Part 1). *Empirical Studies of the Arts, 1,* 17-33.

Hobbs, J.R. (1983b). Metaphor interpretation as selective inferencing: Cognitive processes in understanding metaphor (Part 2). *Empirical Stuidies of the Arts, 1,* 125-141.

Indurkhya, B. (1987). Approximate semantic transference: A computational theory of metaphor. *Cognitive Science, 11,* 445-480.

Johnson, M. (1987). *The body in the mind.* Chicago: Chicago University Press.

Johnson-Laird, J.N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Konolige, K. (1983). A deductive model of belief. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence* (pp. 377-381). Los Altos, CA: Morgan Kaufmann.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: Chicago University Press.

Levesque, H. (1984). A logic of implicit and explicit belief. *Proceedings of the Fourth National Conference on Artificial Intelligence* (pp. 198-202). Los Altos, CA: Morgan Kaufmann.

Maida, A.S. (1983). Knowing intensional individuals. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence* (pp. 382-384). Los Altos, CA: Morgan Kaufmann.

Maida, A.S. (1984). Belief spaces: Foundations of a computational theory of belief (Tech. Rep. No. CS-84-22). University Park: Pennsylvania State University, Department of Computer Science.

Maida, A.S. (1986). Introspection and reasoning about the beliefs of other agents. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 187-195). Hillsdale, NJ: Erlbaum.

Maida, A.S. (1988). A syntactic approach to mental correspondence. *Proceedings of the Canadian Society for Computational Studies of Artificial Intelligence* (pp. 53-58). Edmonton, Alberta.

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

Martins, J., & Shapiro, S. (1988). A model for belief revision. *Artificial Intelligence, 35,* 25-79.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4.* Edinburgh, U.K.: Edinburgh University Press.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill.

Perrault, R. (1987). Invited presentation. *Meeting of the Association for Computational Linguistics.* Stanford, CA.

Perrault, R. (1990). An application of default logic to speech act theory. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Plans and intentions in communication and discourse.* Cambridge, MA: MIT Press.

Perrault, R., & Allen, J. (1980). A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics, 6,* 167-182.

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

Rapaport, W. (1986). Logical foundations for belief representation. *Cognitive Science, 10,* 371-422.

Schiffer, S. (1972). *Meaning.* Oxford, U.K.: Oxford University Press.

Searle, J. (1969). *Speech acts, an essay in the philosophy of language.* Cambridge: Cambridge University Press.

Shadbolt, N. (1983). Processing reference. *Journal of Semantics, 2,* 63-98.

Shapiro, S. (1976). *The relevance of relevance.* Indiana University, Computer Science Department, Bloomington.

Shapiro, S.C., & Rapaport, W.J. (1986). SNePS considered as a fully intensional propositional semantic network. *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 278-283). Los Altos, CA: Morgan Kaufmann.

Sperber, D., & Wilson, D. (1986). *Relevance, communication and cognition.* Oxford, England: Basil Blackwell.

Subramanian, D., & Genesereth, M.R. (1987). The relevance of irrelevance. *Proceedings of the 10th International Joint Conference on Artificial Intelligence* (pp. 416-422). Los Altos, CA: Morgan Kaufmann.

Wiebe, J., & Rapaport, W. (1986). Representing de re and de dicto belief reports in discourse and narrative. *Proceedings of IEEE, 74,* 1405-1413.

Wilks, Y. (1977). Making preferences more active. *Artificial Intelligence, 8,* 75-97.

Wilks, Y. (1986). Relevance and beliefs. In T. Myers, K. Brown, & B. McGonigle (Eds.), *Reasoning and discourse processes.* London: Academic.

Wilks, Y. (1990). Models and theories in artificial intelligence. In D. Partridge & Y. Wilks (Eds.), *The foundations of artificial intelligence: A sourcebook*. Cambridge: Cambridge University Press.

Wilks, Y., & Ballim, A. (1987). Multiple agents and the heuristic ascription of beliefs. *Proceedings of the 10th International Joint Conference on Artificial Intelligence* (pp. 118-124). Los Altos, CA: Morgan Kaufmann.

Wilks, Y., & Ballim, A. (1988). Shifting the belief engine into higher gear. In T. O'Shea & V. Sgurev (Eds.), *Proceedings of the International Conference on AI Methodology and Systems Applications*. Amsterdam: Elsevier Publications.

Wilks, Y., & Ballim, A. (1989a). Belief systems: Ascribing belief. In T. Christaller (Ed.), *Kunstliche intelligenz* (pp. 386-403). Berlin: Springer-Verlag.

Wilks, Y., & Ballim, A. (1989b). The heuristic ascription of belief. In N.E. Sharkey (Ed.), *Models of cognition: A review of cognitive science* (pp. 1-23). Norwood, NJ: Ablex.

Wilks, Y., & Bien, J. (1979). Speech acts and multiple environments. *Proceedings of the Sixth International Joint Conference on Artificial Intelligence* (pp. 968-970). Tokyo, Japan.

Wilks, Y., & Bien, J. (1983). Beliefs, points of view, and multiple environments. *Cognitive Science, 8,* 120-146.

Wilks, T., Fass, D., Guo, C-M., McDonald, J., Plate, T., & Slator, B. (1989). A tractable machine dictionary as a resource for computational semantics. In B.K. Boguraev & T. Briscoe (Eds.), *Computational lexicography for natural language processing*. Harlow, U.K.: Longmans.

Wilks, Y., Fass, D., Guo, C-M., McDonald, J., Plate, T., & Slator, B. (in press). *Providing machine tractable dictionary tools*. In J. Pustejousky (Ed.), *Semantics and the lexicon*. Hingham, MA: Kluwer.