

Your Metaphor or Mine: Belief Ascription and Metaphor Interpretation

Yorick Wilks, John Barnden and Jin Wang
Computing Research Laboratory
New Mexico State University
Box 30001, Las Cruces, N.M. 88003
U. S. A.

Abstract

ViewGen, an algorithm and program for belief ascription, represents the beliefs of agents as explicit, partitioned proposition-sets known as environments. A way of extending *ViewGen* to the interpretation of metaphor, and in particular to the comprehension of metaphor within the belief spaces of particular agents, has been described elsewhere. The paper reports the further refinement and recent implementation of this approach, as well as summarizing the argument for the claim that ordinary non-metaphorical belief ascription and the transfer of information in metaphors can both be seen as different manifestations of a single environment-amalgamation process, one in which explicitly metaphorical amalgamations are triggered by "preference breaking" in the sentence being processed. This requires a consideration of the scoping of metaphor with respect to belief contexts, analogous to the scoping of quantification and definite descriptions with respect to such contexts. As a topic of ongoing and future work, the issue of mixed metaphor, of two distinct types, is briefly addressed.

1 *ViewGen*: The Basic Belief Engine

A computational model of belief ascription is described in detail elsewhere [Wilks and Bien, 1979, 1983] [Ballim, 1987] [Wilks and Ballim, 1987] [Ballim and Wilks, in press] and is embodied in a prolog 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 and Levesque [1985] have also recently explored a belief and speech act logic based on a single explicit default axiom. As our previous work has shown for some years, the default ascription is basically correct, but the phenomena are more complex than are normally captured by an axiomatic approach.

ViewGen also avoids certain counter-intuitive assumptions, such as the *non-persistence* of ignorance about any given proposition p [Perrault, 1990]. Also such systems

avoid any individual-dependent criteria for ascription, such as the individual expertise notions in *Viewgen* (see below).

ViewGen's belief space is divided into a number of explicit, topic-specific partitions, called *topic environments*. *ViewGen* also generates a type of environment known as a *viewpoint*. A viewpoint consists of some person's beliefs about some topics, parcelled up into topic environments. 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

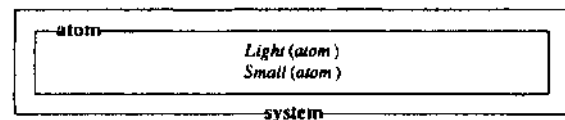


Figure 1: The system's view of an atom.

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. A topic environment contains a group of propositions about the "topic." So, for example, the above diagram conveys that the system believes that atoms are light and small. *ViewGen's* own "knowledge-base" is a viewpoint containing a large number of topic environments.

If the topic of a topic environment is a person, 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, where that will often mean beliefs that conflict 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 would be pictorially represented as in Figure 2. The John

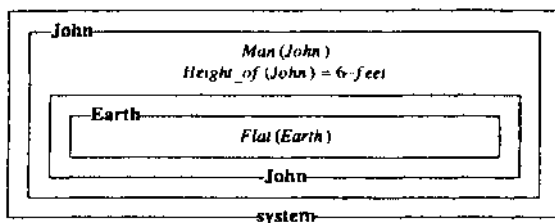


Figure 2: The organization of beliefs about and of John.

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). Environments are dynamically created and altered. The basic algorithm of interest in this paper is 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 of a topic is the same as one's own *except inhere there is explicit evidence to the contrary*. In examples of this sort, where the topic is also the agent into whose environment an ascription is being attempted (i.e., replace "Earth" by "John" in the example), propositions in an outer topic environment E (for the topic John, in the example), are *pushed inwards* into a topic environment (for the same topic) within a believer viewpoint (John's) nested within E. Such inward pushing is central to our later observations on metaphor.

Belief ascription is a far more complex phenomenon than is shown in this brief summary and the key to our method is the delimitation and treatment of cases where the default algorithm is incorrect. 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 and Wilks, in press] [Wilks and Ballim, 1987], and follows a suggestion originally made by McCarthy and Hayes [1969]. This combination of a basic default ascription rule with 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.

2 Metaphor: Shifting the Belief Engine to a Higher Gear

Metaphor is normally explicated, formally or computationally, by a process that transfers properties by some

structural mapping from one structure (the vehicle) to another (the tenor). Classic examples in AI would be the work of Falkenhainer, Forbus and Centner [1989] and of Indurkha [1987]. All these authors are concerned, as we are, with metaphor and analogy viewed as some form of structural mapping; the difference from them of what we offer here is the linkage between that process and the process of belief ascription (and also that of "intensional identification"), as explained in [Ballim, Wilks and Barden, 1991].

We are exploring the application of 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 mentioned above), into an embedded environment for the tenor, to get the tenor seen through the vehicle, or the view of the tenor-as-vehicle.

The key features here are: (1) one of the conceptual domains, namely the metaphor vehicle, is viewed as a "pseudo-believer"; (2) the pseudo-believer has a metaphorical view of a topic or domain; (3) the generation of such a view is not dissimilar from ascribing beliefs to real believers; (4) explicating this by pushing or amalgamating environments yields new intensional entities after an actual transfer of properties.

So, in

Jones threatened Smith's theory by reimplimenting his experiments.

we would know we had a preference-breaking, and potentially metaphorical, situation from the object-feature failure on "threaten" (assuming this expects a person object), at least if we accept the argument of Wilks [1977] that metaphors could be identified, procedurally at least, with the class of preference-breaking utterances. (This includes assertions that violate class relationships, such as "An atom is a billiard ball.")

The awkward cases for that broad delimitation are forms like "Connors 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 use our procedural definition to rule this example out of court as a "garden path metaphor".

We could now plausibly form a metaphoric view of theory-as-person using the environment-amalgamation process sketched above. Figure 3 shows possible system environments for theory and person, and the resulting theory-as-person environment, where the arrow indicates the new environment resulting from the application of the default rule when the properties of the outer environment (for Person) are amalgamated into the inner environment (for Theory) and survive unless contradicted (as the Concrete predicate in fact is). So, 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

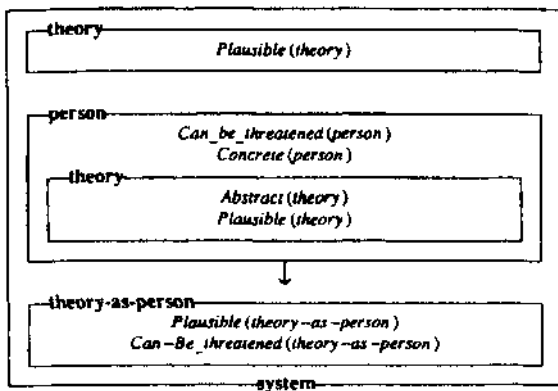


Figure 3: Forming a theory-as-person environment.

produce no different a set of plausible properties transferred than any other system (e.g. that of Falkenhainer *et al.*); and that would be, again, an experimental question, as would be the very difficult issue how many "irrelevant" properties are ascribed as well, and whether there is a danger of them swamping the metaphorically relevant ones. But the importance or originality we propose 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, as we argued in detail in [Ballim *et al.*, 1991].

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 that the content of belief is often 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 no help at all. Are the entities referred to somehow metaphorically the Prussian people, etc., 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"; perform a trial pushdown 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, that there is no space to discuss here, would be that we examined our inventory of representations to see if we had one for "Prussia" that already expressed the (dead) metaphor of a country-name-as-apology (some would insist that this was a metonymy, but we do not address this alternative here).

The amalgamation of the notions of belief ascription and metaphor interpretation is described and justified in detail in [Ballim *et al.*, 1991]. The method as described there explicitly addresses the case of metaphor only at the top level of discourse, within the system's own notions of the component elements of the metaphor. But of course, in the general process of metaphor interpretation, we must relativize these processes to the space of any relevant believer. The system might believe the zoological truism that pigs are clean and neat, but when talking to the non-zoologist John, use the standard metaphor "He's a pig" on the assumption that John would perform the ascription against the dirty-unhygienic-pig belief that the system believes John to hold. The system must be able to model *that* metaphorical ascription, whose elements the system itself does not believe.

This relativization is included in a recent implementation of the approach within *VtwGen* by Ballim, as extended by one of us (Wang). The need for relativization underscores the benefits of our method of unifying metaphorical transference with belief ascription. We shall analyze such examples later in more detail, but we must first discuss an issue of "metaphorical scope" that is implicitly raised by the above.

3 Metaphorical Scope

Consider the sentence

John believes that a cure for terrorism is needed.

The complement of this belief report — the clause following the word "that" — can be construed as involving a terrorism-as-disease metaphor. We must first realize that there is a *metaphor scoping* issue here. The "inner scope" reading involves the idea that John himself thinks of terrorism as a disease (and we might predict that he would report his belief by means of the sentence "A cure for terrorism is needed"). The "outer scope" reading is that John believes something about terrorism that is being portrayed *by the speaker* in terms of disease-curing. John does not necessarily have a belief couched in these terms, nor would he necessarily report his belief in these terms. (Perhaps John would say: "Something needs to be done to eliminate terrorism and repair the damage it has done to society.")

We now concentrate on the inner scope reading. In the process of setting up a topic environment for terrorism inside John's belief environment, the system needs to perform (i) the normal default ascription process, here moving some of its own beliefs about terrorism down into John, *as well as* performing (ii) a metaphoric transference process. This combination of tasks is simplified by our method of dressing the tasks in essentially the same algorithmic clothes.

This is especially so since it is *John's* view of disease that is important in the metaphoric transference, not the system's. After all, John might believe that diseases are caused by demonic influences, say, and this belief could affect what he would think of as reasonable ways of curing terrorism — e.g., exorcism. Thus, in task (ii), a disease topic environment must be set up within John's

The treatment is such that it makes no difference whether (i) A-as-B is seen as C or (ii) A is seen as B-as-C. This seems to accord well with our intuitions about examples we have considered. We plan to tackle parallel mixing as well in future work, especially as both types are common in the realm of metaphors of mind (which form the central concern of Barnden [1989, 1990]).

Acknowledgments

This paper builds on work done in conjunction with Afzal Ballim. We are grateful also for useful suggestions from Subhankar Banerjee, David Fawell and Dihong Qiu.

References

- [Ballim, 1987] Afzal Ballim. The subjective ascription of belief to agents. In *Advances in artificial intelligence*. Edited by J. Hallam and C. Mellish. Chichester, England: John Wiley & Sons.
- [Ballim and Wilks, in press] Afzal Ballim and Yorick Wilks. *Artificial Believers*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- [Ballim et al., 1991] A. Ballim, Y. Wilks and J. Barnden. Belief ascription, metaphor, and intensional identification. *Cognitive Science*, 15, pp. 133-171.
- [Barnden, 1989] John A. Barnden. Belief, metaphorically speaking. In *Proceedings of the 1st Intl. Conf on Principles of knowledge Representation and Reasoning*. San Mateo, CA: Morgan Kaufman.
- [Barnden, 1990] John A. Barnden. Naïve Metaphysics: a metaphor-based approach to propositional attitude representation (Unabridged version). *Memoranda in Computer and Cognitive Science*, MCCS-90-174. Computing Research Laboratory, New Mexico State University.
- [Bochenski, 1961] I. Bochenski. *A History of Formal Logic*. Indiana: Notre Dame University Press.
- [Cohen and Margalit, 1972] J. Cohen and A. Margalit. The role of inductive reasoning in the interpretation of metaphor. In *Semantics of natural language*. Edited by D. Davidson and G. Harman. Dordrecht: Reidel.
- [Cohen and Levesque, 1985] P. R. Cohen and H. J. Levesque. Speech acts and rationality. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*. University of Chicago.
- [Davidson, 1978] D. Davidson. What metaphors mean. *Critical Inquiry*, 5, pp.31-48.
- [Falkenhainer et al., 1989] B. Falkenhainer, K. Forbus and D. Gentner. The structure-mapping engine: algorithm and examples. *Artificial Intelligence*, 41, pp.1-63.
- [Indurkha, 1987] B. Indurkha. Approximate Semantic Transference: A Computational Theory of Metaphor. *Cognitive Science*, 11, pp.445-480.
- [Lakoff, 1971] G. Lakoff. On generative semantics. In *Semantics*. Edited by D. Steinberg & L. Jakobovits. London and New York: Cambridge University Press.
- [McCarthy and Hayes, 1969] J. McCarthy and P. Hayes. Some Philosophical problems from the Standpoint of Artificial Intelligence. In *Machine Intelligence*. Edited by Neltzer & Mitchie. Edinburgh University Press.
- [Perrault, 1987] R. Perrault. Invited presentation. *Meeting of the Association for Computational Linguistics*. Stanford, Calif.
- [Perrault, 1990] R. Perrault. An application of default logic to speech act theory. In *Plans and intentions in communication and discourse*. Edited by P. Cohen, J. Morgan & M. Pollack. Cambridge, Mass.: MIT Press.
- [Rieger, 1975] C. Rieger. Conceptual Memory and Influence. Edited by R. Schank. North Holland: Amsterdam.
- [Wilks, 1975] Y. Wilks. Preference Semantics. In *Formal Semantics of Natural Language*. Edited by E. Keenan. Cambridge, UK: Cambridge University Press.
- [Wilks, 1977] Y. Wilks. Making preferences more active. *Artificial Intelligence*, 8, pp.75-97.
- [Wilks and Ballim, 1987] Y. Wilks and A. Ballim. Multiple agents and the heuristic ascription of belief. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*. Los Altos, Calif.: Morgan Kaufmann.
- [Wilks and Bien, 1979] Y. Wilks and J. Bien. Speech acts and multiple environments. In *Proceedings of the 6th International Joint Conference on Artificial Intelligence*. Tokyo, Japan.
- [Wilks and Bien, 1983] Y. Wilks and J. Bien. Beliefs, points of view and multiple environments. In *Cognitive Science*, 8, pp.120-146