

PROPOSITIONAL ATTITUDES, COMMONSENSE REASONING, AND METAPHOR

John A. Barnden

Computing Research Laboratory
New Mexico State University

THE CLAIM TO BE ARGUED

People have highly metaphorical ways of regarding each others' minds, as is well known and will be exemplified later. This means that in practice people will often adopt metaphorical views of each others' propositional attitudes (their states of belief, hope, expectation, intention, desire, and so on). This has been ignored in detailed technical studies of how to represent propositional attitudes in AI systems. Yet, this paper will argue that it is absolutely crucial to pay attention to such views in designing propositional attitude representation schemes.

The paper briefly argues this claim about metaphor, which I have made recently elsewhere [Barnden 1988], but also places it in the context of the following more general claim:

In the design of representation schemes for beliefs and other propositional attitudes, it is essential to pay attention to issues of commonsense, plausible reasoning (including metaphorical reasoning as a special case), as opposed to the strict deductive reasoning that is traditionally considered."

We shall refer to this as The Claim. There are, certainly, some well-recognized connections between propositional attitudes and commonsense, plausible reasoning, and these links have already affected attitude-representation design to some extent. However, the paper will press for a new effect on such design. Also, the *strongest* previous effects on attitude-representation design have come not from commonsense reasoning but from deductive reasoning (see next section). This accounts for the contrast to deductive reasoning in The Claim.

The plan of the paper is as follows. This section will briefly review some existing links between propositional attitudes and commonsense reasoning. The next section will unify important aspects of the traditional interplay of deductive reasoning and propositional attitudes under a certain banner (the "inference-lifting question schema"), but will argue that this interplay is just a special and somewhat artificial special case of a more general interplay that allows for commonsense reasoning also. Then, three sections will lead up to showing the importance in this interplay of *metaphorical* reasoning, because of people's metaphorical views of attitudes. We will see that attitudes and commonsense reasoning are more intricately intertwined than is commonly realized — to the extent that commonsense plausible reasoning must be given a bigger role in the design of genuinely useful propositional-attitude representation techniques than it has had in the past. This role is at last sketched in the penultimate section.

As for recognized connections between attitudes and commonsense reasoning, the following ones are among the most important. A central aspect of discourse and speech act theory is listeners' plausible conclusions about the intentions, beliefs, etc. of speakers. [See e.g. Allen 1983, Grosz & Sidner 1986, Sperber & Wilson 1986]. Automated tutoring systems should ideally make plausible inferences about the beliefs, expectations and intentions of users. Another connection is shown in some research that has been done on resource-limited belief reasoning [e.g. Konolige 1983; see also Fagin & Halpern 1987]. Commonsense reasoning is involved in choosing between different types of reading of propositional attitude reports, e.g. between "de-re" and "de-dicto" readings [Wiebe & Rapaport 1986]. There has recently been research on mapping between plausible-reasoning logics and belief logics [Konolige 1987]. Maida (1986) has worked on introspection-based, plausible ascription of belief to other agents. Issues of plausible, default-based ascription of beliefs have been addressed by Ballim (1987), Wilks & Ballim (1987), and Barnden & Ballim (1988) argue that the issues have a (not well-recognized) effect on attitude-representation design.

COMMONSENSE REASONING AND ATTITUDES: INTRODUCTION

The Claim concerns the issue of *how inference requirements affect detailed representational design* for propositional attitude purposes. Most relevant work within AI and philosophy has been on

- (a) respecting the referential opacity of propositional-attitude contexts;
- (b) preventing "logically omniscient" inference [see e.g. Levesque 1984, Fagin & Halpern 1987];
- (c) while, in some research, allowing simple types of logical combination or analysis of beliefs, as in inferring to and/or fro between "Z believes that R and S" on the one hand and "Z believes that R" and "Z believes that S" on the other.

Issues (a) to (c) are mainly concerned with the following question schema:

IL: The Inference-Lifting Question-Schema

GIVEN that the statement Q follows, in sense F, from some statements P,

DOES the statement "Z believes (hopes, etc.) that Q" follow, in sense G, from the statements "Z believes (hopes, etc.) that P"?

The term "statement" is used in a loose, general sense in this paper, covering both natural language sentences and formal representational expressions.

The following will assume that, in the schema, A's propositional attitude is one of belief (by far the most common case discussed). In one important version of issue (a), there are two P statements, e.g. "Jim's wife is clever" and "Mary is Jim's wife"; Q is derived from the first of these by co-referential substitution, giving "Mary is clever"; and senses F and G are both deductive consequence. The answer to the question is "no" when "Jim's wife" in the statement "Z believes that Jim's wife is clever" is interpreted in a de-dicto way. Note also that the referential opacity issue is also taken to cover the case when Mary is Jim's wife but agent Z is *not* presumed to believe this.

In issue (b), the P statements are a statement R and a statement to the effect that R implies Q, and senses F and G are again deductive consequence. Logical omniscience is commonly taken to include other slightly different patterns of deductive inference about agent Z [see e.g. Fagin & Halpern 1987], but these are less germane to this paper.

Thus we see that issues (a), (b) and, obviously, (c) are primarily concerned with F and G being *definite* (i.e. deductive) inference, as opposed to *plausible, commonsense* inference, even though the issues are indeed genuinely concerned with commonsense inference in the indirect or suppressed sense that the proscriptive thrust of (a) and (b), and the permissive thrust of (c), are inspired by commonsense observations *by the theorist* on what it is reasonable to infer about ordinary agents.

However, just as most inference of true practical interest in AI and cognitive science generally is of a commonsensical and merely plausible nature (based on defaults, abduction, metaphor, and so on), the instances of IL that are of *primary* interest are those in which sense F and/or sense G brings in plausible, commonsense reasoning. One such commonsense instance of IL has A's attitude being belief, Q being "Peter can fly", the P being the statements "Usually birds can fly" and "Peter is a bird", and F being the sense of default reasoning. G is also likely to be some plausible-reasoning sense, although notice that the degree of plausibility in the inference to "Z believes that Q" from the "Z believes that P" has no necessary, simple connection with the degree of plausibility that would obtain if the system were to infer Q from the P. For example, the system may happen to know that John almost always does in fact makes (at least tentative) inferences about the flying ability of particular birds, so that the inference to "John believes that Q" could have a much higher plausibility than the system's potential inference to Q from the P.

The confidence inequality can be the other way around: Q might even follow *definitely* from the P, as far as the system is concerned, and yet the system may only have limited confidence in the

conclusion that Z believes Q. Thus even the traditional issue of IL instances where F is a definite-reasoning sense, *G should be taken to be a plausible-reasoning sense*.

I have not explored any implications that IL instances such as the one about birds might have for attitude-representation design. We proceed now to build a bridge towards IL instances that do have such implications.

COMMONSENSE INSTANCES OF THE IL SCHEMA: "EXPLICATIONS"

Consider the attitude report: *Professor Z believes that X's theory is threatened*.

An AI system interpreting this sentence may well have to *explicate* the notion of threat involved here, in order to adequately link the information conveyed by the sentence with other information. For instance, suppose that the system believes that Professor Z is usually correct in her beliefs about whether there are experiments providing evidence against given theories. Suppose also that one reasonable explication of the notion of a theory being threatened (out of perhaps many possible explications available to the system) is that *there is an experiment that provides evidence against the theory*. Then, if the system considers it reasonable to impute this explicated view of theory-threat to Z, Z can use the attitude report to make the plausible inference that

Professor Z believes that there is an experiment providing evidence against X's theory.

The system can go on to use its belief about Z's reliability on such matters to make the plausible inference that there is indeed an experiment providing evidence against X's theory.

We have here an instance of the IL schema in which P is the statement that X's theory is threatened, and Q is the statement that there is some experiment providing evidence against X's theory. It is clear that G is some sense of *plausible inference*. As for sense F, we could suppose that Q follows from P in the sense of being a plausible *analysis* of the meaning of P. Or, we could suppose that Q follows from P in the sense of being an ordinary plausible *inference* from P. This sort of distinction actually makes little difference to the concerns of this paper.

In the example we started with the natural language sentence displayed above, but there is no implication that the conclusion (Prof. Z believes that there is some experiment ...) is also a natural language sentence. Rather, we may take the system to be inferring from the given sentence to some internal representational expression for the conclusion. We are here unconventionally subsuming the natural language understanding process under inference. This subsumption is mainly for the sake of simplicity of discussion, and it makes little ultimate difference to the issues of concern in this paper.

It is potentially a complex matter for the system to devise a reasonable explication of a vague notion, like that of a theory being threatened, embedded in a belief. In our example, the explication can depend on the academic fields of X and Z (and these fields may differ), default information about how people in Z's field view theories in X's field and their relationship to experiments, special information about Z's view of these matters, and so on. Also, there can be explicit hints about specific explications, as in "Professor Z believes that X's theory is threatened by Y's more economical theory". Here an experiment-based explication that might otherwise have been used would be discarded in favor of an explication in terms of opposing theories. The complexities in the choice of reasonable explications are beyond the scope of this paper but are the subject of active research by the author.

METAPHORICAL INSTANCES OF THE IL SCHEMA

So far we have ignored the possible metaphorical nature of the threat notion. Let us now assume that the notion is metaphorical, and that the vehicle of the metaphor is a WAR domain [see e.g. Lakoff & Johnson 1980] in which people can threaten, battle with, and conquer each other, and so on.

It is then important for the AI system to be able to relate the example sentence of the last section to statements such as the following, which stay within the same metaphor:

Prof. Z believes that Y intends to attack X's theory,

given that this sentence is interpreted to mean that the attacking aspect of the WAR metaphor *is in fact being used mentally by Z*. (This reading is analogous to de-dicto interpretation of referential phrases in belief contexts.) If one of the reasonable explications of "threatened" that is available to the system is to the effect that someone is intending to attack the threatened thing, then the system could come to the plausible conclusion that the threat to X's theory believed by Z is in fact just Y's intention (according to Z) to attack it. Notice that this conclusion rests not only on the plausibility of the indicated explication of theory-threat within the indicated metaphor, but also on the plausibility of ascribing this metaphorical view to Z.

Very similar issues arise with an example sentence like

General Z believes that terrorism is creeping across the globe.

Suppose the system has been told that "The belief that terrorism is slowly and stealthily moving across the globe is excessively pessimistic". The system should be able to come to the plausible conclusion that Z is being excessively pessimistic. What is at issue here is, again, commonsense inference *within* a particular metaphor. This type of inference serves for sense F in the IL schema. Sense G is some form of plausible inference. P is the statement that terrorism is creeping across the globe, and Q the statement that terrorism is slowly and stealthily moving across the globe. It seems implausible and uneconomical to suggest that the AI system should connect the two metaphorical statements about terrorism by going through some literal explications. Probably no adequate ones are available anyway.

Thus, the situations with which this section began, in which a word like "threatened" indicates a metaphor, and in which a metaphorical explication of the word's import should be constructed, are just a special case of a more general metaphor-based inferential phenomenon subsumed by the IL schema. However, it is the special case that leads us into the crux of the argument, as follows.

METAPHORICAL EXPLICATIONS OF ATTITUDES

The notions within attitude contexts that need to be given possibly-metaphorical explications include *propositional attitudes themselves*. Consider the sentence

Z hopes that X will come to realize that his theory is faulty.

We will see that the propositional-attitude notion of realizing may have to be treated analogously to the notion of being threatened — and, in particular, that it is likely that the notion of realizing will have to be given some *metaphorical* explication that is suggested by discourse context. The argument is a condensation of the more detailed discussion in Barnden (1988). During the argument it is important to bear in mind that what is at issue is the explications and metaphors that *the agent Z* can be reasonably taken to have in mind for the notion of realizing.

Commonsense, metaphorical views of mind that people entertain have received some close attention [see e.g. Lakoff & Johnson 1980, Larsen 1987, Reddy 1979, Tomlinson 1986]. Now, suppose the above sentence is succeeded by the following one, which brings in the metaphor of X's mind as a battleground in which forces (ideas, habits, etc.) engage in struggles:-

But Z is afraid that the idea of the theory being faulty is having an uphill struggle against X's habit of wishful thinking.

Assume that it is reasonable to take Z herself (not just the speaker) to be thinking in terms of the metaphor. Then the most commonsensically reasonable and coherent way to approach the two sentences is to explicate, in terms of that metaphor, the realization by X that is hoped for by Z in the first sentence. That is, the system should, say, take Z to hope that *the idea of X's theory being faulty will gain dominance in the battleground of X's mind*.

It might be argued that instead of *introducing* a metaphor into the interpretation of the first sentence, the system should seek to *eliminate* the metaphorical aspect of the second sentence. This argument, however, faces the well-known difficulty of eliminating metaphors for mental states/processes [see e.g. Fainsilber & Ortony 1987]. Also, the mind-as-battleground metaphor might be continued and further elaborated in succeeding sentences, such as:

Z has had much experience of conflicts of this sort in her colleagues.

In sum, what is needed is a system that can explicitly cast an agent, such as Z in the example, as using a particular metaphor, or mix of metaphors, chosen from a range of possible metaphors for propositional attitudes (the attitude of realizing, in the example). Notice also that the chosen metaphor(s) may differ from default ones the system might be tempted to use for a realizing that does not appear within another attitude. The system might, for instance, use by default the metaphor of realizing as being a matter of solidifying an idea into concrete form.

THE EFFECT ON ATTITUDE REPRESENTATION

When the AI system meets the sentence ‘‘Z hopes that X realizes that his theory is faulty’’ and interprets it, the interpretation might well use some standard, atomic way of representing the state of realizing, as for instance in the modal-logic formula

hopes(Z, realizes(X, faulty(theory-of(X))))).

Thus, there is a limited sense in which the representation of attitudes is not affected by the considerations of the previous section. However, these considerations show that in order for the internal, formal rendering of the sentence to do useful work in relation to later sentences (or, of course, earlier ones), it will typically have to be replaced, or at least supplemented, by a rendering in which X’s realization is explicated in appropriate metaphorical terms. As a simplified illustration, the new rendering could be:

hopes(Z, will-dominate(idea-of(X, faulty(theory-of(X))), battleground-in(mind-of(X))))).

The thesis is that *this* is the sort of representation (of what it is that Z reportedly hopes for) that will be heuristically useful.

Notice that the effect on representation that has been claimed comes down to an effect on the sort of predicate symbols (for instance) that are used, on the assumed ontology underlying the representation, and on the degree of elaborateness of attitude representations. The arguments have not displayed a reason for thinking that traditional *styles* of representation — e.g. modal, quotational, concept-based, situation-based¹ — are inadequate. However, my previous work on attitude representation [Barnden 1983, 1986a,b, 1987a,b] shows that the use of quotational, concept-based and situation-based styles has a natural tendency to lead to a *commitment to a particular explication* of an attitude that is nested inside and attitude of an agent Z, where, moreover, the explication is typically a *commonsensically implausible* one to impute to Z — whereas, of course, we have argued for the use of any of a *variety of commonsensically plausible* explications to impute to Z. Commitments to particular explications can be entirely avoided, but only at a significant cost of complexity in the representational syntax (see Barnden (1987a,b) for the quotational case). It is, on the other hand, relatively easy to replace the explications the styles tend to enforce by commonsensically plausible ones, using formulae analogous to the second one displayed in the previous section. Thus, the present paper’s considerations do impinge upon the broad style of representation, to the extent of recommending that if the mentioned styles of representation should only be used with caution and in a particular way.

¹ See e.g.: for modal style: Chellas 1980, Halpern & Moses 1985; for quotational style: Haas 1986, Perlis 1985, Quine 1981; for concept-based style: McCarthy 1979, Creary 1979, Creary & Pollard 1985; for situation-based style: Barwise & Perry 1983, Hobbs 1985.

The modal-logic style was not addressed in the previous paragraph, but there are considerations mitigating against its usefulness in practical contexts, including: the often-noted inability to quantify over beliefs and treat them as individuals; and the inability to deal with (metaphorical) belief explications that are explicitly signalled, as in "Z believes that some experiment, of whose identity she has only a hazy idea, is a threat to X's theory" [Barnden 1987a,b]. This sentence suggests that a belief is composed of ideas which can be hazy.

CONCLUSION

We saw that a consideration of the metaphorical explications of propositional attitudes nested within other attitudes has a fundamental effect on how we should approach the task of designing a system that can represent attitudes so as to reason about them adequately in practical situations.

The issue of metaphorical attitude-explications is within the intersection of two more general issues: that of metaphors within attitude contexts, and that of commonsensical explications of notions (e.g. "threatened") within attitude contexts. These two issues are in turn about special commonsense cases of the Inference-Lifting schema IL, and illuminate the much more general point that the instances of schema IL that are of primary interest are those where senses F and/or G bring in *commonsense*, *plausible* reasoning, as opposed to deductive reasoning as is traditional.

A detailed representational approach based on the ideas of this paper is being developed from the quotational-cum-concept-based scheme described in Barnden (1987a,b).

ACKNOWLEDGMENTS

I am grateful for encouragement, stimulation, and constructive criticism from Afzal Ballim, Dan Fass, Sylvia Candelaria del Ram, and Yorick Wilks.

REFERENCES

- Allen, J. (1983). Recognizing intentions from natural language utterances. In M. Brady & R.C. Berwick (Eds), *Computational models of discourse*, Cambridge, Mass.: MIT Press.
- Ballim, A. (1987). The subjective ascription of belief to agents. In J. Hallam & C. Mellish (Eds), *Advances in artificial intelligence*. Chichester, UK: Ellis Horwood.
- Barnden, J. A. (1983). Intensions as such: an outline. *Procs. 8th Int. Joint Conf. on Artificial Intelligence*, Karlsruhe, W. Germany.
- Barnden, J.A. (1986a). Imputations and explications: representational problems in treatments of propositional attitudes. *Cognitive Science*, 10 (3), 319-364.
- Barnden, J.A. (1986b). A viewpoint distinction in the representation of propositional attitudes. In *Proceedings of the 5th Nat. Conf. on Artificial Intelligence (AAAI86)*, Philadelphia, Penn.
- 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*, Elsevier (North-Holland): Amsterdam.
- Barnden, J.A. (1987b). Avoiding some unwarranted entailments among nested attitude reports. *Memoranda in Computer and Cognitive Science*, No. MCCS-87-113, Computing Research Laboratory, New Mexico State University.
- Barnden, J.A. (1988). Belief, metaphorically speaking. Submitted to 1988 European Conf. on Artificial Intelligence (ECAI-88).
- Barnden, J.A. & Ballim, A. (1988). The relevance of belief environments. Submitted to 7th National Conf. on Artificial Intelligence (AAAI-88).
- Barwise, J. & Perry, J. (1983). *Situations and attitudes*. Cambridge, Mass.: MIT Press.

- Chellas, B.F. (1980). *Modal logic*. Cambridge University Press.
- Creary, L. G. (1979). Propositional attitudes: Fregean representation and simulative reasoning. *Procs. 6th. Int. Joint Conf. on Artificial Intelligence*, Tokyo.
- Creary, L.G. & Pollard, C.J. (1985). A computational semantics for natural language. *Procs. 23rd Ann. Meeting of the Association for Computational Linguistics*, Univ. of Chicago.
- Fagin, R. & Halpern, Y.J. (1987). Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34 (1), 39-76.
- Fausilber, L. & Ortony, A. (1987). Metaphorical uses of language in the expression of emotions. *Metaphor and Symbolic Activity*, 2 (4), 239-250.
- Grosz, B.J. & Sidner, C.L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12 (3), 175-204.
- Haas, A.R. (1986). A syntactic theory of belief and action. *Artificial Intelligence*, 28, 245-292.
- Halpern, J. Y. (ed.) (1986). *Theoretical aspects of reasoning about knowledge: proceedings of the 1986 conference*. Los Altos, CA: Morgan Kaufmann.
- Halpern, J.Y. & Moses, Y.O. (1985). A guide to the modal logics of knowledge and belief: preliminary draft. In *Procs. 9th. Int. Joint Conf. on Artificial Intelligence*, Los Angeles, 1985.
- Hobbs, J.R. (1985). Ontological promiscuity. *Procs. 23rd Ann. Meeting of the Association for Computational Linguistics*, Univ. of Chicago.
- Konolige, K. (1983). A deductive model of belief. *Procs. 8th Int. Joint Conf. on Artificial Intelligence*, Karlsruhe, W. Germany.
- Konolige, K. (1987). On the relation between default theories and autoepistemic logic. *Procs. 8th Int. Joint Conf. on Artificial Intelligence*, Karlsruhe, W. Germany.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: U. Chicago Press.
- Larsen, S. F. (1987). Remembering and the archaeology metaphor. *Metaphor and Symbolic Activity*, 2 (3), 187-199.
- Levesque, H. J. (1984). A logic of implicit and explicit belief. *Procs. Natl. Conf. on Artificial Intelligence*, Univ. of Texas at Austin.
- Maida, A. S. (1986). Introspection and reasoning about the beliefs of other agents. *Procs. 8th Conference of the Cognitive Science Society*.
- McCarthy, J. (1979). First order theories of individual concepts and propositions. In J. E. Hayes, D. Michie & L. I. Mikulich (Eds.), *Machine Intelligence 9*. Chichester: Ellis Horwood.
- Pertlis, D. (1985). Languages with self-reference I: Foundations. *Artificial Intelligence*, 25, 301-322.
- Quine, W.V.O. (1981). Intensions revisited. In W.V. Quine, *Theories and things*. Cambridge, Mass: Harvard U. Press.
- Reddy, M. J. (1979). The conduit metaphor — a case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and Thought*, Cambridge, UK: Cambridge University Press.
- Sperber, D. & Wilson, D. (1986). *Relevance: communication and cognition*. Oxford: Basil Blackwell.
- Tomlinson, B. (1986). Cooking, mining, gardening, hunting: metaphorical stories writers tell about their composing processes. *Metaphor and Symbolic Activity*, 1 (1), 57-79.
- Wiebe, J.M. & Rapaport, W.J. (1986). Representing *de-re* and *de-dicto* belief reports in discourse and narrative. *Procs. of the IEEE*, 74 (10), 1405-1413.
- Wilks, Y. & Ballim, A. (1987). Multiple agents and the heuristic ascription of belief. *Procs. 10th Int. Joint Conf. on Artificial Intelligence*, Milan, Italy.