MODALITY, SI! MODAL LOGIC, NO!

John McCarthy
Computer Science Department
Stanford University
Stanford, CA 94305
jmc@cs.stanford.edu
http://www-formal.stanford.edu/jmc/

1997 Mar 18, 5:23 p.m.

Abstract

This article is oriented toward the use of modality in artificial intelligence (AI). An agent must reason about what it or other agents know, believe, want, intend or owe. Referentially opaque modalities are needed and must be formalized correctly. Unfortunately, modal logics seem too limited for many important purposes. This article contains examples of uses of modality for which modal logic seems inadequate.

I have no proof that modal logic is inadequate, so I hope modal logicians will take the examples as challenges.

Maybe this article will also have philosophical and mathematical logical interest.

Here are the main considerations.

Many modalities: Natural language often uses several modalities in a single sentence, *"I want him to believe that I know he has lied."* [Gab96] introduces a formalism for combining modalities, but I don't know whether it can handle the examples mentioned in this article.

New Modalities: Human practice sometimes introduces new modalities on an *ad hoc* basis. The institution of owing money or the obligations the

Bill of Rights imposes on the U.S. Government are not matters of basic logic. Introducing new modalities should involve no more fuss than introducing a new predicate. In particular, human-level AI requires that programs be able to introduce modalities when this is appropriate, e.g. have function taking modalities as values.

Knowing what: "Pat knows Mike's telephone number" is a simple example. In [McC79b], this is formalized as

$$knows(pat, Telephone(Mike)),$$

where *pat* stands for the person Pat, $Mike$ stands for a standard concept of the person Mike and $Telephone$ takes a concept of a person into a concept of his telephone number. We might have

$$telephone(mike) = telephone(mary),$$

expressing the fact that Mike and Mary have the same telephone, but we won't have

$$Telephone(Mike) = Telephone(Mary),$$

which would assert that the concept of Mike's telephone number is the same as that of Mary's telephone number. This permits us to have

$$\neg knows(pat, Telephone(Mary)).$$

even though Pat knows Mike's telephone number which happens to be the same as Mary's. The theory in [McC79b] also includes functions from some kinds of things, e.g. numbers or people, to standard concepts of them. This permits saying that Kepler did not know that the number of planets is composite while saying that Kepler knew that the number we know to be the number of planets (9) is composite.

The point of this example is not mainly to advertise [McC79b] but to advocate that a theory of knowledge must treat *knowing what* as well as *knowing that* and to illustrate some of the capabilities needed for adequately using *knowing what.*

Presumably

$$knows(pat, Telephone(Mike))$$

could be avoided by writing

$$(\exists x)(knows(pat, Telephone(Mike) = x)),$$

but the required "quantifying in" is likely to be a nuisance.

Proving Non-knowledge [McC78] formalizes two puzzles whose solution requires inferring non-knowledge from previously asserted non-knowledge and from limiting what is learned when a person hears some information.[1]

[McC78] uses a variant of the Kripke accessibility relation, but here it is used directly in first order logic rather than to give semantics to a modal logic. The relation is $A(w1, w2, person, time)$ interpreted as asserting that in world $w1$, it is possible for $person$ that the world is $w2$. Non-knowledge of a term in $w1$ is e.g. the color of a spot or the value of a numerical variable, is expressed by saying that there is a world $w2$ in which the value of the term differs from its value in $w1$.

[Lev90] uses a modality whose interpretation is *"all I know is ....".* He uses autoepistemic logic [Moo85], a nonmonotonic modal logic. This seems inadequate in general, because we need to be able to express *"All*

[1] The *three wise men puzzle* is as follows:

A certain king wishes to test his three wise men. He arranges them in a circle so that they can see and hear each other and tells them that he will put a white or black spot on each of their foreheads but that at least one spot will be white. In fact all three spots are white. He then repeatedly asks them, "Do you know the color of your spot?" What do they answer?

The solution is that they answer, *"No,"* the first two times the question is asked and answer *"Yes"* thereafter.

This is a variant form of the puzzle which avoids having wise men reason about how fast their colleagues reason.

Here is the *Mr. S and Mr. P* puzzle:

Two numbers m and n are chosen such that $2 \leq m \leq n \leq 99$. Mr. S is told their sum and Mr. P is told their product. The following dialogue ensues: Mr. P: I don't know the numbers.

Mr. S: I knew you didn't know. I don't know either.

Mr. P: Now I know the numbers.

Mr S: Now I know them too.

In view of the above dialogue, what are the numbers?

I know about the value of x is . . . .” [2] Here’s an example. At one stage in *Mr. S and Mr. P*, we can say that all Mr. P knows about the value of the pair is their product and the fact that their sum is not the sum of two primes.

[KPH91] treats the question of showing how President Bush could reason that he didn’t know whether Gorbachev was standing or sitting and how Bush could also reason that Gorbachev didn’t know whether Bush was standing or sitting. The treatment does not use modal logic but rather a variant of circumscription called autocircumscription proposed by Perlis [Per88].

Joint knowledge and learning In the wise men problem, they learn at each stage that the others don’t know the colors of their spots, and in Mr. S and Mr. P they learn what the others have said. In each case the learning is joint knowledge, wherein several people knowing something jointly implies not only that each knows it but also that they know it jointly. [McC78] treats joint knowledge by introducing pseudo-persons for each subset of the real knowers. The pseudo-person knows what the subset knows jointly. The logical treatment of joint knowledge in [McC78] makes the joint knowers S5 in their knowledge. I don’t know whether a more subtle axiomatization would avoid this.

[McC78] treats learning a fact by using the time argument of the accessibility relation. After $person$ learns a fact p the worlds that are possible for him are those worlds that were previously possible for him and in which p holds. Learning the value of a term is treated similarly.

Other modalities [McC79a] treats believing and intending and [McC96] treats introspection by robots. Neither paper introduces enough formalism to provide a direct challenge to modal logic, but it seems to me that the problems are even harder than those previously treated.

Acknowledgements: This work was supported in part by DARPA (ONR) grant N00014-94-1-0775. Tom Costello provided some useful discussion.

[2] Halpern and Lakemeyer in [HL95] show that the quantified version of Levesque’s logic is incomplete, but this is a different complaint from the one we make here.

References

[Gab96] Dov Gabbay. Fibred semantics and the weaving of logics: Part I: Modal and intuitionistic logics. *Journal of Symbolic Logic*, 61(4):1057–1120, 1996.

[HL95] Joseph Y. Halpern and Gerhard Lakemeyer. Levesque's axiomatization of only knowing is incomplete. *Artificial Intelligence*, 74(2):381–387, 1995.

[KPH91] Sarit Kraus, Donald Perlis, and John Horty. Reasoning about ignorance: A note on the Bush-Gorbachev problem. *Fundamenta Informatica*, XV:325–332, 1991.

[Lev90] Hector J. Levesque. All I know: a study in autoepistemic logic. *Artificial Intelligence*, 42:263–309, 1990.

[McC78] John McCarthy. Formalization of two puzzles involving knowledge[3], 1978. Reprinted in [McC90].

[McC79a] John McCarthy. Ascribing mental qualities to machines[4]. In Martin Ringle, editor, *Philosophical Perspectives in Artificial Intelligence*. Harvester Press, 1979. Reprinted in [McC90].

[McC79b] John McCarthy. First Order Theories of Individual Concepts and Propositions[5]. In Donald Michie, editor, *Machine Intelligence*, volume 9. Edinburgh University Press, Edinburgh, 1979. Reprinted in [McC90].

[McC90] John McCarthy. *Formalization of common sense, papers by John McCarthy edited by V. Lifschitz.* Ablex, 1990.

[McC96] John McCarthy. Making Robots Conscious of their Mental States[6]. In Stephen Muggleton, editor, *Machine Intelligence 15.* Oxford University Press, 1996.

[3] http://www-formal.stanford.edu/jmc/puzzles.html

[4] http://www-formal.stanford.edu/jmc/ascribing.html

[5] http://www-formal.stanford.edu/jmc/concepts.html

[6] http://www-formal.stanford.edu/jmc/consciousness.html

[Moo85] Robert C. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25(1):75–94, January 1985.

[Per88] Donald Perlis. Autocircumscription. *Artificial Intelligence*, 36:223–236, 1988.

/@steam.stanford.edu:/u/jmc/w97/modality1.tex: begun Sat Mar 1 11:21:20 1997, latexed March 18, 1997 at 5:23 p.m.