MODALITY FOR ROBOTS—RESPONSES TO HALPERN AND WANSING

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

1999 Sep 15, 3:02 p.m.

Abstract

(McCarthy 1997) has attracted responses defending modal logic from Heinrich Wansing (Wansing 1998) and Joseph Halpern (Halpern 1999). My criticism of modal logic in connection with AI is that modal logic, at least as described in the literature, isn't expressive enough for an independently operating robot. It relegates to humans reasoning with and about modalities that an independent robot will have to do for itself.

The demand wasn't sufficiently clearly expressed in (McCarthy 1997), and perhaps consequently the responses don't sufficiently speak to it.

Heinrich Wansing (Wansing 1998) and Joseph Halpern (Halpern 1999) responded to (McCarthy 1997) which argued that modal logic was inadequate to meet the requirements for the treatment of modality in AI and that formalizations of modality in coventional logic, e.g. first order logic, have greater potential. Wansing and Halpern make many similar points, but the ones I want to address are stated by Halpern in a more convenient form, so I'll concentrate on them. I have included more formulas and more explanations of formulas than in (McCarthy 1997). Here are some considerations.

1. Joseph Halpern (Halpern 1999) writes *"McCarthy has argued that modal logic is too limited for various purposes. I consider the extent to which he is right."* I evidently did not make clear that my primary purpose is to make a language suitable for representation of facts by a robot acting independently with human level capability. This means that information often regarded as metalinguistic must also be available to the robot. If possible worlds are important, the robot must reason about possible worlds. For this modal logic is inadequate.

2. I agree with Halpern that the possible-worlds structure can sometimes help illuminate arguments. (McCarthy 1978) uses possible worlds explicitly in the formal argument. On the other hand (Kraus et al. 1991) infers non-knowledge using second order logic but neither modal logic nor possible worlds.

3. I agree that the decidability of propositional modal formalisms is useful, and robots should be able to use the decision procedures. The procedures can also be applied to first order axiomatizations of the same modalities. (Grädel et al. 1997) relates this to the decideability of two variable first order logic.

4. I don't agree with Halpern's statement that common, i.e. joint, knowledge is not expressible in first order logic. Halpern's statement depends on regarding common knowledge as a transitive closure of iterated knowledge of the several knowers and requiring that the formalization of this transitive closure be complete. This is not the best way to handle common knowledge. (McCarthy 1978) treats common knowledge by introducing virtual persons possessing the common knowledge of a finite set of real persons (two or three persons in the examples of that paper.) Since transitive closure is not completely formalizable in first order logic, common knowledge as in my paper will not have all the properties of transitive closure. However, it does have enough of the properties to do the problems of the wise men and of Mr. S and Mr. P.

   I don't know how to say whether that notion is adequate for other uses of common knowledge in common sense reasoning. The notion of common knowledge in that paper satisfies S5, and I now regard that as a blemish to be fixed. The reason is that while the S5 property of common knowledge is adequate for the problems treated in the paper, it would make inconsistent a more powerful system that includes Peano arithmetic (or elementary syntax, to use the terminology of (Montague 1963)). I want robots to be able to reason with ZFC, which subsumes Peano arithmetic and to be able to assume that other robots also know ZFC.

5. I propose that a robot be able to introduce new modalities as new predicates. In logicians' terminology, this changes the language but not the logic. I have regarded introducing new modalities to modal logical systems as changing the logic. Perhaps these are similar ideas, but some interactive theorem provers for logic, e.g. Jussi Ketonen's EKL, allow operations that define new predicates. I don't know whether any modal verification systems allow the introduction of new modal operators in the course of a proof. (Costello and Patterson 1998) gives a system where all modal operators that could be defined in first order logic can be defined by three new operators introduced in the system.

6. *Knowing what*. Halpern considers that putting *knowing what* directly in the language is a "convoluted approach" and prefers treating *knowing what* as a satellite of *knowing that*. I think treating *knowing what* directly as $knows(pat, Telephone(Mike))$ corresponds more closely to natural language usage than does Halpern's

$$\exists x K_{Pat}(\textit{telephone-\#}(Mike) = x).$$

It is merely not what modal logicians are used to. This corresponds to the fact that in English, *"Pat knows Mike's telephone number"* is more natural than *"There is a number concept X such that Pat knows that Mike's telephone number is X"*. In the language of (McCarthy 1979), this would be

$$\begin{aligned}&(\exists x)(\textit{phone-number}(x)\\ &\wedge k(pat, Equal(Telephone(Mike), Concept1(x))).\end{aligned} \tag{1}$$

We have to make sure that Pat knows that Mike's telephone number is a certain number. In our system

$$(\exists X)(k(pat, Equal(Telephone(Mike), X)) \tag{2}$$

is always true, since we can substitute $Telephone(Mike)$ for X and get $k(pat, Equal(Telephone(Mike), Telephone(Mike)))$, which will be an instance of a general theorem.

To avoid making Halpern's $(\exists x)K_{Pat}(telephone\text{-}\#(Mike) = x)$ a tautology, the system of modal logic must have *non-rigid designators*, i.e. constants that take on different values in different possible worlds, and there must be a restriction on instantiation of bound variables to rigid designators. Rigid designators have been controversial in philosophical logic and presumably have disadvantages.

Writing K_{Pat} suggests that the knower argument of K is not ever intended to be a variable over which we quantify, and Joe Halpern confirms this. But *"Nobody knows the troubles I've seen"* illustrates that quantification over knowers is common in ordinary language and the resulting sentences can themselves be the objects of knowledge, and we want among the facilities for robot use of modality. Quite possibly quantifying over knowers can be added to modal logic, but the decidability results for modal logic may not extend to such formulas.

I wrote $concept1(x)$, to emphasize that other functions from objects to concepts of them may be useful, e.g. $concept2(x)$. (McCarthy 1979) gives additional examples of how treating concepts as objects gives flexibility.

Here are two of the examples.

$$\neg knew(kepler, Composite(Number(Planets))). \tag{3}$$

In (3), $Planets$ is a concept of the set of planets, $Number(Planets)$ is a concept of the number of planets, and $Composite(Number(Planets))$ is the proposition that this number is composite. Note that capital letters are used for concepts and for functions from concepts to concepts. *kepler* denotes the person Kepler and not some concept of him, and $knew(kepler(\ldots))$ asserts that Kepler knew something. Since Kepler presumably thought the number of planets was 7, he presumably did not know that the number of planets is composite.

$$knew(kepler, Composite(Concept1(denot(Number(Planets))))), \tag{4}$$

which use functions *denot* from a concept to the thing it denotes and $Concept1$ going from a thing to a standard concept of it, both of these being partial functions. Assuming that the number of planets is 10, this expresses the fact that Kepler knew that this number is composite.

The following sentence attributed to Russell is discussed by Kaplan: *"I thought that your yacht was longer than it is"*. We can write it

$$\begin{aligned}&believed(i, Greater(Length(Youryacht)),\\&Concept1(denot(Length(Youryacht))))\end{aligned}. \tag{5}$$

where we are not analyzing the pronouns or the tense, but are using *denot* to get the actual length of the yacht and $Concept1$ to get back a concept of this true length so as to end up with a proposition that the length of the yacht is greater than that number.

If we introduce $belief(i, X)$ to denote what I believe the numerical value of the denotation of the concept X, to be, we can write

$$belief(i, Length(Youryacht)) > length(youryacht), \tag{6}$$

which is more straightforward but probably yet harder than the previous entities to express in modal form. The first part of the equation answers the question "What did I think was the length of your yacht?", which might have the answer "40 feet".

(McCarthy 1979) argues for using propositions and individual concepts rather than strings of letters on the grounds that the same concept may be denoted by different strings of letters, e.g. we may want "P and Q" and "Q and P" to name the same proposition.

Informal language distinguishes between concepts of objects and objects themselves. Trying to avoid the distinction in formal languages limits what can be expressed. (Frege 1892) and (Church 1951) make these distinctions.

7. (McCarthy 1978), first actually published in (McCarthy 1990), treats non-knowledge by formalizing possible worlds in first order logic, using an extended Kripke accessibility relation. $A(w1, w2, person, time)$ means that world $w2$ is accessible from world $w1$ for $person$ at $time$. Putting in time permits including the effects of learning in the formalism. Thus the worlds accessible at time $t+1$ comprise the subset those worlds accessible at time t in which the proposition learned is true.

This approach using possible worlds expresses non-knowledge of the value of an expression by asserting the existence of possible worlds in which the expression has different values. For example, in the puzzle of Mr. S and Mr. P we are told that initially Mr. S knows only the sum of the two numbers.[1] We have

$$(\forall pair)(sum(pair) = Sum0 \rightarrow (\exists w)(A(RW, w, MrS, 0) \wedge pairfun(w) = pair)). \tag{7}$$

Here RW denotes the real world[2], the quantification is over pairs of numbers, $sum(pair)$ is the sum of the numbers of the pair, $Sum0$ is the sum told to Mr. S, RW is the real world, and $pairfun(w)$ is the pair of numbers associated with the world w.

Actually, we need another level of knowledge in order to say that everyone knows Mr. S knows only the sum. This makes the formula

$$\begin{array}{l}(\forall rw)(A(RW, rw, joint(MrS, MrP), 0) \rightarrow \\ \quad (\forall pair)(sum(pair) = Sum0 \rightarrow \\ \qquad (\exists w)(A(rw, w, MrS, 0) \wedge pairfun(w) = pair))).\end{array} \tag{8}$$

[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?

[2] —the best of all possible worlds—

Here $joint(MrS, MrP)$ is the pseudo-person who has the joint knowledge of Mr. S and Mr. P. The occurrence of the real world RW in (7) is replaced by the variable rw.

If we understand the problem of Mr. S and Mr. P in terms of possible worlds, so should the robot. Actually, it would be better to use something less grandiose than the full Stalnaker-Lewis notion of possible world. Better would be possible worlds limited to the a context, e.g. one associated with the Mr. S and Mr. P puzzle.

8. Halpern includes *"In particular, although people have tried to capture notions like intentions and desires using possible world, I am not convinced that it is the best way to go; possible worlds is certainly not the answer for all problems"*. What concerns me about this sentence is the possible implication that there will always be a person around to decide what formalism to use. Our formalism for modality needs to be expressive enough, so that we can imagine the robot deciding for itself what formalism to use for a problem.

9. I haven't yet been able to make a logical language including modality that has the full capabilities that I consider needed for independently thinking robots. However, it seems to require the ability to express the change of what is known after learning, knowing what, allowing proofs of non-knowledge and joint knowledge of groups of actors. It may also have to be able to express facts about modalities as objects.

Summary and challenges.

1. Do the "Kepler knew ..." and "Your yacht ..." examples.

2. What about functions from objects to concepts of them?

3. How is a robot to reason with metalinguistic information, e.g. to reason about possible world structures?

4. What about quantifying over knowers?

My general opinion is that keeping the mathematical structure of modal logic has interfered with making it useful in AI and for applications, e.g. to databases.

Acknowledgments:

I'm indebted to Tom Costello and Pat Hayes for useful discussions.

This research was partly supported by U.S. Air Force Office of Scientific Research contract #F49620-97-1-0207 under the AFOSR New World Vistas program and by the Defense Advanced Research Project Agency High Performance Knowledge Bases Program.

References

Church, A. 1951. A formulation of the logic of sense and denotation. In P. Henle (Ed.), *Essays in honor of Henry Sheffer*, 3–24. New York.

Costello, T., and A. Patterson. 1998. Quantifiers and Operations on Modalities and Contexts. In *Proceedings of Sixth Intl. Conference on Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann.

Frege, G. 1892. Uber sinn und bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik* 100:25–50. Translated by H. Feigl under the title "On Sense and Nominatum" in H. Feigl and W. Sellars (eds.) *Readings in Philosophical Analysis*. New York 1949. Translated by M. Black under the title "On Sense and Reference" in P. Geach and M. Black, *Translations from the Philosophical Writings of Gottlob Frege*. Oxford, 1952.

Grädel, E., P. G. Kolaitis, and M. Y. Vardi. 1997. On the decison problem for two variable first-order logic. *Bulletin of Symbolic Logic* 3(1):53–69.

Halpern, J. Y. 1999. On the adequacy of modal logic. in ETAI discussion.

Kraus, S., D. Perlis, and J. Horty. 1991. Reasoning about ignorance: A note on the Bush-Gorbachev problem. *Fundamenta Informatica* XV:325–332.

McCarthy, J. 1978. Formalization of two puzzles involving knowledge[3]. Reprinted in (McCarthy 1990).

McCarthy, J. 1979. First Order Theories of Individual Concepts and Propositions[4]. In D. Michie (Ed.), *Machine Intelligence*, Vol. 9. Edinburgh: Edinburgh University Press. Reprinted in (McCarthy 1990).

McCarthy, J. 1990. *Formalizing Common Sense: Papers by John McCarthy*. 355 Chestnut Street, Norwood, NJ 07648: Ablex Publishing Corporation.

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

McCarthy, J. 1997. Modality, si! modal logic, no! *Studia Logica* 59:29–32.

Montague, R. 1963. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica* 16:153–167. Reprinted in (Montague 1974).

Montague, R. 1974. *Formal Philosophy.* Yale University Press.

Wansing, H. 1998. Modality, of course! modal logic, si! *Journal of Logic, Language and Information* 7(3):iii–vii.