On the Adequacy of Modal Logic

Joseph Y. Halpern*
Cornell University
Computer Science Department
Ithaca, NY 14853
halpern@cs.cornell.edu
http://www.cs.cornell.edu/home/halpern

July 19, 1999

Abstract

McCarthy has argued that modal logic is too limited for various purposes. I consider the extent to which he is right.

McCarthy [1997] has argued that modal logic is inadequate for many purposes. There are certainly situations for which modal logic is inadequate, just as there are situations for which first-order logic is in adequate. My goal in this brief note is to look more carefully at the relative adequacy of modal logic and first-order logic, in particular with respect to some of the issues raised by McCarthy in his article.

First, we'd better define some terms, at least informally. By "modality", I mean notions like "knowing", "believing", and "intending". By "modal logic", I mean a logic that includes modalities as syntactic operators (so that it allows formulas like $K_3\varphi$—"agent 3 knows φ") and gives semantics to these operators using possible worlds.

There seems to be general agreement (at least, among the three participants in this debate, including me) that modalities are necessary to express the types of things we want to express in AI. The question is whether the appropriate logical formalism to express them using modal logic. My answer (as usual for questions such as these) is "it depends". Modal logic is a useful tool, but it doesn't solve all problems. The same, of course, can be said for first-order logic. It seems to me that sometimes we are better off using propositional modal logic, sometimes we are better off using first-order logic, sometimes we need first-order modal logic (i.e., both first-order quantification, modal operators, and possible worlds), and sometimes not even first-order modal logic is adequate.

When is propositional modal logic better than first order logic? Here are three cases.

*Supported in part by NSF under grant IRI-96-25901.

- The possible-worlds structure can sometimes help illuminate arguments. Consider the *muddy children puzzle* [Fagin, Halpern, Moses, and Vardi 1995; Halpern and Moses 1990] (which is essentially isomorphic to the cheating wives puzzle or the wise-men-with-spots-on-their-head puzzle). A father comes along and sees children playing in the mud. He says "Some of you have mud on your foreheads." Then he keeps asking the children whether they know they have mud on their own foreheads. The children can see the foreheads of all the other children but not their own. Nevertheless, after the father's third question, each of the (three) children with mud on their forehead says "Yes, I know I have mud on my forehead." A possible-worlds analysis of this puzzle, using binary relations between worlds to represent what worlds a child considers possible, shows that the structure corresponding to the puzzle is an n-dimensional cube, which gets truncated each time the father speaks (see [Fagin, Halpern, Moses, and Vardi 1995; Halpern and Vardi 1991]). Such a graphical description of the puzzle can help clarify what is going on in ways that first-order logic cannot do. In general, the graphical structure of a possible-worlds model makes seems to help the reasoning process.

- The satisfiability/validity problem for propositional modal is decidable (typically PSPACE-complete or EXPTIME-complete, depending on the logic involved [Halpern and Moses 1992]). First-order logic is undecidable. Thus, to the extent that we are interested in automating reasoning, there are certainly advantages to using propositional modal logic (at least, if it is adequate for expressing what we want).

- There are notions that can be expressed using propositional modal logic that cannot be expressed in first-order. One example is *common knowledge*—everyone knows that everyone knows that everyone knows .... Common knowledge has been used in AI (indeed, it was first introduced by McCarthy [McCarthy, Sato, Hayashi, and Igarishi 1979]), economics [Aumann 1976], and distributed computing [Halpern and Moses 1990]. There are standard translations from propositional modal logics of knowledge to first-order logic going back to van Benthem [1974, 1985] that capture the modal operators K_i by a binary relation. The binary relation corresponding to common knowledge is the transitive closure of the union of the relations corresponding to the K_i operators. However, an easy argument using compactness shows that transitive closure is not expressible in first-order logic. Thus, this approach cannot be used to capture common knowledge.

There are clearly times when we need the expressive power of first-order quantification. Whether it is better in these cases to use just first-order logic or first-order modal logic depends, as I said before, on the details of the application. Making the decision clearly involves understanding the limitations of both. While modal logic certainly has its limitations, many of the issues raised by McCarthy [1997] can in fact be dealt with perfectly adequately. Others are indeed serious problems, but seem to be equally difficult to handle in first-order logic (or any other flavor of logic that I'm aware of). Here are more detailed comments on the specific issues raised by McCarthy.

Many modalities: There is no question that we often need to use multiple modalities. Although the combination of modalities sometimes raises interesting philosophical and technical questions, it is invariably a straightforward matter to at least give semantics to multiple modalities if the semantics for the individual modalities are well understood. To give just two examples, Kraus and Lehmann [1988] deal with knowledge, belief, and time, and Fagin and Halpern [1994] deal with knowledge for many agents, common knowledge, and probability. There are many other examples in the literature.

New modalities: I agree that with McCarthy that "Human practice sometimes introduces new modalities on an *ad hoc* basis." No logic that I am aware of can handle this, at least not in full generality. Logic in general starts with a fixed syntax—it can't be changed on the fly. Of course, in first-order logic, we can always add relation symbols to a vocabulary. In the same way, we can add new modal operators to modal logic. I'm not sure if this counts as a solution to McCarthy's concern though. In any case, it seems to me that modal logic fares no better or worse than any other approach in terms of dealing with new modalities.

Knowing what: McCarthy [1997] discusses a problem first raised in [McCarthy 1979] involving knowledge. We want to model a situation where Pat knows Mike's telephone, Mike's telephone number is the same as Mary's telephone number, but Pat does not know Mary's telephone number. He suggests what seems to me a convoluted approach to modeling this that involves distinguishing the *concept* of a person or a telephone number from the actual person or telephone number. Here is a case where possible worlds helps clarify what is going on, without needing to introduce concepts in this sense. In first-order modal logic, the fact that Pat knows Mike's telephone number is expressed as $\exists x K_{Pat}(\mathit{telephone\text{-}\#(Mike)} = x)$. This means that there is a number x such that in all the worlds that Pat considers possible, Mike's telephone number is x. The fact that Mike's telephone number is the same as Mary's telephone number is expressed as *telephone-#(Mike) = telephone-#(Mary)*. This simply says that in the actual world, the two telephone numbers are the same. It does not force them to be the same in all the worlds that Pat considers possible. (This would be expressed by the formula $K_{Pat}(\mathit{telephone\text{-}\#(Mike)} = \mathit{telephone\text{-}\#(Mary)})$.) As a consequence, it does not follow that Pat knows Mary's telephone number, since there may be a world that Pat considers possible where Mary's telephone number is different from Mike's. That is, in first-order modal logic, the following formula is *not* valid:

$$(\mathit{telephone\text{-}\#(Mike)} = \mathit{telephone\text{-}\#(Mary)}) \wedge \exists x K_{Pat}(\mathit{telephone\text{-}\#(Mike)} = x) \Rightarrow \exists x K_{Pat}(\mathit{telephone\text{-}\#(Mary)} = x).$$

The fact that $t_1 = t_2$ in the actual world for some terms t_1 and t_2 does not allow us to substitute t_2 for t_1 in a modal context, and for good reason! McCarthy [1997] mentions concerns with quantifying-in raised by the existential quantifier outside

the scope of the modal operator, but this in fact is not a problem; see [Grove 1995] for further discussion of this issue.

Proving non-knowledge: Dealing with non-knowledge or "only knowing" is hard, especially in a multi-agent setting. (See, for example, [Halpern and Lakemeyer 1996] for some discussion of this issue.) The problem is equally difficult for modal logic and first-order logic (and, I believe, will continue to be difficult even if we go to higher-order logics). The real problem lies in giving good semantics to notions such as "all I know", whatever approach we end up using. My guess is that model-theoretic reasoning is a better way to deal with here than developing a formal logic, but this remains an open problem.

Joint knowledge and learning: It wasn't clear to me exactly what McCarthy meant by "joint knowledge" among the agents in a group G; it could be "common knowledge among the agents in G", "everyone in group G knows" or "if the agents in group G pooled their knowledge together they would know" (this is called *distributed knowledge* [Fagin, Halpern, Moses, and Vardi 1995; Halpern and Moses 1992]). All of these notions can be captured in modal logic. The assumptions made about the underlying notion of knowledge determine the properties of joint knowledge. For example, common knowledge satisfies S5 if the knowledge of the agents does. However, it does not necessarily satisfy KD45 if the underlying notion of knowledge does.

Regarding learning, this can certainly be captured (at least to some extent) in models of knowledge and time; see, for example, [Halpern and Vardi 1989].

Other modalities: I am not convinced that modal logic (that is, possible worlds) is necessarily the best way to capture all modalities. In particular, although people have tried to capture notions like intentions and desires using possible worlds, I am not convinced that it is the best way to go; possible worlds is certainly not the answer for all problems. On the other hand, I do not know any better formalization of these notions in first-order logic.

As I hope this note makes clear, I don't think that modal logic is a universal panacea; neither is first-order logic. Choosing the "right" logic to use in a given application is still an art, which requires a detailed understanding of the domain and the underlying logics. All I have tried to do in this brief note is to clarify some of the issues that need to be considered in making the choice.

References

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics 4*(6), 1236–1239.

van Benthem, J. F. A. K. (1974). Some correspondence results in modal logic. Report 74–05, University of Amsterdam.

van Benthem, J. F. A. K. (1985). *Modal Logic and Classical Logic*. Naples: Bibliopolis.

Fagin, R. and J. Y. Halpern (1994). Reasoning about knowledge and probability. *Journal of the ACM 41*(2), 340–367.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi (1995). *Reasoning about Knowledge*. Cambridge, Mass.: MIT Press.

Grove, A. J. (1995). Naming and identity in epistemic logic II: a first-order logic for naming. *Artificial Intelligence 74*(2), 311–350.

Halpern, J. Y. and G. Lakemeyer (1996). Multi-agent only knowing. In Y. Shoham (Ed.), *Theoretical Aspects of Rationality and Knowledge: Proc. Sixth Conference*. San Francisco, Calif.: Morgan Kaufmann.

Halpern, J. Y. and Y. Moses (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM 37*(3), 549–587. A preliminary version appeared in *Proc. 3rd ACM Symposium on Principles of Distributed Computing*, 1984.

Halpern, J. Y. and Y. Moses (1992). A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence 54*, 319–379.

Halpern, J. Y. and M. Y. Vardi (1989). The complexity of reasoning about knowledge and time, I: lower bounds. *Journal of Computer and System Sciences 38*(1), 195–237.

Halpern, J. Y. and M. Y. Vardi (1991). Model checking vs. theorem proving: a manifesto. In J. A. Allen, R. Fikes, and E. Sandewall (Eds.), *Principles of Knowledge Representation and Reasoning: Proc. Second International Conference (KR '91)*, pp. 325–334. San Francisco, Calif.: Morgan Kaufmann. An expanded version appears in *Artificial Intelligence and Mathematical Theory of Computation (Papers in Honor of John McCarthy)* (ed. V. Lifschitz), Academic Press, 1991, pp. 151–176.

Kraus, S. and D. Lehmann (1988). Knowledge, belief, and time. *Theoretical Computer Science 58*, 155–174.

McCarthy, J. (1979). First-order theories of individual concepts and propositions. In D. Michie (Ed.), *Machine Intelligence*, Volume 9. Edinburgh: Edinburgh University Press.

McCarthy, J. (1997). Modality, si! Modal logic, no! *Studia Logica 59*, 29–32. Also available at http://www-formal.stanford.edu/jmc/modality/modality.html.

McCarthy, J., M. Sato, T. Hayashi, and S. Igarishi (1979). On the model theory of knowledge. Technical Report STAN-CS-78-657, Stanford University.