CS-TR-3278 | May 26, 1994
UMIACS-TR-94-62

LOGIC FOR A LIFETIME

Don Perlis

Institute for Advanced Computer Studies
Computer Science Department
AV Williams Bldg
University of Maryland
College Park, MD 20742

perlis@cs.umd.edu

Abstract

There has been an explosion of formal work in commonsense reasoning in the past fifteen years, but almost no significant connection with work in building commonsense reasoning systems (cognitive or otherwise). We explore the reasons, and especially the ideal formal assumption of omniscience, reviewing and extending arguments that this is irreparably out of line with the needs of any real reasoning agent. On the other hand, this exploration reveals some desiderata that might still be given useful formal treatment, but with a somewhat altered set of aims from what has motivated most formal work. The discussion is motivated by several examples of commonsense reasoning, involving change of belief in addition to the more usual arguments concerning resource limitations. Key to the entire discussion is the notion that real reasoners do not usually have the luxury of isolated problems with well-defined beginnings and endings, but rather must deal with evolving and ongoing problems and situations.

This research was supported in part by NSF grant IRI9311988.

LOGIC FOR A LIFETIME

Donald Perlis
Institute for Advanced Computer Studies
and
Department of Computer Science
University of Maryland
College Park, MD 20742
(301) 405-2685
(301) 405-6707
perlis@cs.umd.edu

Abstract

There has been an explosion of formal work in commonsense reasoning in the past fifteen years, but almost no significant connection with work in building commonsense reasoning systems (cognitive or otherwise). We explore the reasons, and especially the ideal formal assumption of omniscience, reviewing and extending arguments that this is irreparably out of line with the needs of any real reasoning agent. On the other hand, this exploration reveals some desiderata that might still be given useful formal treatment, but with a somewhat altered set of aims from what has motivated most formal work. The discussion is motivated by several examples of commonsense reasoning, involving change of belief in addition to the more usual arguments concerning resource limitations. Key to the entire discussion is the notion that real reasoners do not usually have the luxury of isolated problems with well-defined beginnings and endings, but rather must deal with evolving and ongoing problems and situations.

Areas: reasoning;belief-change;contradiction;omniscience;resource-limitations

LOGIC FOR A LIFETIME

Donald Perlis

Institute for Advanced Computer Studies

and

Department of Computer Science

University of Maryland

College Park, MD 20742

1 Introduction

There has been an explosion of formal work in commonsense reasoning in the past fifteen years, largely in the specific area of nonmonotonic reasoning (NMR).[1] This resulted in part from the observation [17] that human commonsense reasoning (CSR) often does not obey traditional modes of logical inference. But Minsky may have misdiagnosed the source of the problem. He is right that traditional (monotonic) logic fails to model CSR, but I will argue that this is not so much due to an inherent nonmonotonicity in CSR as it is to the *omniscience* of traditional logic: all formulas that can be proven are in fact proven (made into theorems)—or, in model-theoretic terms, all semantic consequences of one's axioms are believed. Omniscience prevents proper treatment of change in belief; this theme will be elaborated in later sections.

The problem of omniscience has not gone unnoticed by formalists. There is a standard attitude toward this, what I will call the *standard model*, a justification of the formal approaches despite the known inappropriateness of omniscience. At a very high level (a more prosaic and more revealing description is given later) it is this: omniscient formalisms have the major advantage of being simpler and easier to study, and can be taken as modeling

[1] E.g., witness the collections [6] and [8], the recurrent international NMR and KRR workshops, and in particular the many beautiful discoveries by McCarthy, Reiter, Moore, Konolige, Levesque, Pearl, Halpern, and Lifschitz, among others.

ideal reasoners against which real (human or robotic) reasoners can be measured as approximations. The invitation to analogy with ideal gas laws and real gases is strong: we do learn useful things about real gases from ideal gas models; in many situations a real gas behaves a lot like an ideal gas. Whether a similar useful relation is obtains between ideal (omniscient) and real commonsense reasoners is the topic of this paper.

There appear to be several pieces of evidence that this research tradition might not relevantly address the issues facing the design of a real commonsense reasoner, not even in useful approximation, and that omniscience is irreparably out of line with the needs of any real reasoning agent.[2] Here I present and discuss these pieces of evidence, as well as their possible significance for future formal directions, since this exploration also suggests desiderata that may very well be given useful formal treatment, but with a somewhat different set of aims from what has motivated much existing formal work.[3] To a considerable extent, the paradigm suggested by Nilsson [18] of a robot with a lifetime of its own serves as an underlying motivational theme.

2 The standard model

We begin with Minsky's (by now famous and overworked) examples [17] of two commonsense human inferences: from the information that Tweety is a bird, one may well infer that Tweety can fly; and yet if instead the reasoner had originally had the additional information that Tweety is an ostrich, the former inference would likely not have occurred and indeed instead one may have inferred that Tweety cannot fly. Thus more information may actually block a conclusion. This so (by now, at least) so obvious as to be a totally unsurprising observation about human behavior, and by extension about intelligent robot behavior. But the clear conclusion is that traditional monotonic logic is not the proper vehicle for much of (human

[2]This may account for the fact that those building commonsense reasoning systems (e.g., [21, 24, 22, 23, 10, 27, 28, 7]) have availed themselves of only modest borrowings from traditional NMR formalisms.

[3]Thus this is not at all an anti-logicist essay, but rather a call for yet further improved formalisms. The NMR revolution of the 1980s was a real step forward in the liberation of logic from traditional settings and toward greater realism about the nature of commonsense reasoning. We may now be in need of yet another revolution.

or robot) commonsense reasoning.

By 1980 at least three distinct formalisms for NMR had been developed [12, 25, 14], and the standard model began to emerge. To present this model, we first restate the examples in chronological terms: at first we know Tweety is a bird and so conclude Tweety can fly; later we learn Tweety is an ostrich, and so then retract our earlier conclusion. According to the standard (nonmonotonic) model of reasoning (a folklore view that evolved in the early 1980s, but to my knowledge has never been carefully expressed in writing), commonsense reasoning consists of an ongoing alternation of two kinds of symbolic manipulation: the CSR phase, during which defeasible theorems are proven from given commonsense axioms (beliefs), and truth-maintenance phase (TMS, see [2]), during which the axiom set is updated based on new incoming information (and theorems are retracted as needed). Then follows another round of CSR, then (if more data comes in) more TMS, etc.

In the CSR phase, the reasoner's beliefs are precisely the set of *all*[4] theorems (or semantic consequences) emanating from the commonsense axioms (whatever the notion of proof or consequence is). The kind of nonmonotonic effort by which the beliefs are produced is not generally examined; nor is the precise nature of the TMS update phase. But the formal relation between the original belief (theorem/consequence) set and the new (post-update) belief set is given close attention for therein lies the nonmonotonicity and the judgement as to whether the appropriate "reasoning" has taken place. Thus in the case of Tweety, in phase 1 (see Figure 1 below) the reasoner believes Tweety can fly, since this follows nonmonotonically from the axiom that Tweety is a bird; in phase 1', the information that Tweety is an ostrich is supplied as a new axiom and the belief that Tweety can fly is dropped, readying us for phase 2 in which now the reasoner believes (this time perhaps from ordinary monotonic logic and the background knowledge that ostriches cannot fly) that Tweety cannot fly. The mechanisms of belief change are not of interest in the standard model, nor even the TMS phase which is generally not explicitly mentioned; rather only the formal relationship between phase 1 and

[4]This is the omniscience: whatever follows is believed. Thus Fermat's Last Theorem is believed (if we can believe Andrew Wiles!); and if we believe a contradiction then we believe everything whatsoever since everything follows from that.

2, between 2 and 3, etc, is of interest.

```
***************************************************************************
phase:  1             1'             2             2'             3


       CSR     |      TMS     |     CSR     |      TMS     |     CSR     | ...
  -------------?-------------|-------------?-------------|-------------?--
    belief set |update axioms| new beliefs |update axioms| new beliefs | ...
                     etc                          etc


                                   -------------------> time


  (ignore the ?s for now)
****************************** Figure 1 **********************************
```

The time taken to reason (i.e., the time spent in the CSR phases) can be ignored (all one's beliefs are instantaneously ready-to-hand); and inference (reasoning) shuts down during the TMS phase which merely inspects the proof trees to see what no longer has justification under the new axioms. In effect, the TMS phase transforms one theory (CSR belief set) into a new one. Thus the course of nonmonotonic reasoning is seen as a succession of theories, each fixed and perfect for its role as given by its associated axioms.

The standard objection: resource limitations The usual (word-of-mouth) objection to this model is that it is doubly impossible. Not only is it impossible for a real reasoner to have an infinite set of beliefs (as is usually required) but due to the nonmonotonicity, the beliefs are not in general even computable from the axioms. Moreover, it takes time (and space) to produce beliefs (theorems). Finally, from a contradiction we (people) do not come to believe everything; we either do not notice the contradiction or we do and take remedial action.

The standard rejoinder: approximation The usual rejoinder is that the standard model is an idealization, that real reasoners can be seen as approximations to the ideal model, and either (i) as technology produces faster computers the distinction will, for practical problems, fade away, or (ii) the distinction will remain a large but useful measure for comparing one robot to another in terms of which comes closer to the ideal. And contradictory beliefs are unusual occurrences, not part of ordinary everyday reasoning.

This quarrel can be pursued further; but I leave it here because I want to aim at a rather different set of objections to the standard model.

3 The standard model revisited

Looking again at the figure above, we see ?s in the separations between phases 1 and 1', between 2 and 2', and so on. These are to call our attention to these very crucial interfaces. Somehow the logic engine that produces defeasible beliefs in the CSR phase must cease doing so when new axioms come in, so that TMS can take place. Now since the standard model supposes CSR to be instantaneous, this is not a conceptual problem, and for a real (e.g., human) reasoner, we can suppose that new data simply shuts off other trains of thought for a moment. But now comes a difficulty.

What is an axiom? How does a reasoner decide that new data is to be taken as axiomatic, trusted over other data? Aside from logical truths, what do we know for certain? Or how do we prioritize our beliefs in order of believability? We clearly do, at least to some extent, since we often give up some beliefs in favor of others. However, some examples will show that this is far from trivial.

Suppose I watch the TV meteorologist in front of all her weather maps, saying that last night the temperature reached a low of one degree below zero, Fahrenheit. This is an expert opinion about an already measured datum, and is accepted by me without any apparent inferential steps. Then my four-year old son says that Bill Clinton is six feet 8 inches tall, and I reject this, thinking that (i) my son often exaggerates and (ii) if Clinton were that tall this surely would be frequently mentioned in the news and I would have heard about it again and again (yet I have never heard it except from my son). My belief that Clinton is less than six feet eight is clearly nonmonotonic (autoepistemic, to be precise) and my one-degree-below-zero belief is less clearly so. If the latter is to be considered defeasible, then which of our beliefs is not defeasible?

Yet I would not be startled to learn that the meteorologist misread the temperature from her notes, or that the thermomenter was broken, and that in fact the temperature last night reached a low of only three degrees above zero. This is not such a shocking development. But it would be far more shocking, disturbing to my sense of how things work, to learn that Clinton is in fact six-eight.

So, it appears that little indeed is axiomatic. When new data comes in, do we trust it? We go through some complicated reasoning, including assessments of how information about Presidents is reported, about how easily we remember things, and so on. That is, we use substantial portions of our commonsense world view: we do commonsense reasoning to help assess whether to trust what we hear. So, we cannot turn the CSR inference engine off while we attach new axioms: we must keep the engine running.

This is particularly the case when we are presented with contradictory data. Thus if we hear Tweety is an ostrich but we have already seen Tweety flying, we are not so quick to do the Minskian switch. We tend to trust our own eyes (Tweety is flying); but not always (maybe that bird is not Tweety). While there may well be formal priority principles here, if so then they depend richly on the fabric of our overall world view and so cannot be relegated to a TMS phase in which CSR is turned off: dealing with conflicting data is part and parcel of what commonsense reasoning is all about.

Finally, every time TMS is called for in the standard model, there is a case of contradiction of a previous belief and a new datum. Thus contradictions are as common as is change of belief: it is contradictions that signal us that a change is needed, that it is time to rethink our thinking.

4 Dealing with contradictions

How can a reasoning apparatus (person, robot, program) deal with contradictions? There have been various proposals. Some, such as the paraconsistent logics surveyed in [1], aim to extract a trustworthy core of inferences while avoiding the contradictions. Others, such as [5, 15, 26], aim to detect and resolve contradictions. The latter are closer in spirit to the needs we are addressing here.

Unlike the traditional view that abhors a contradiction and seeks at all costs to avoid such[5] and fears that CSR will come to naught (or to disaster) in their presence, the "new" view being presented here is that contradictions are our friends, guiding us to look more closely at what we are thinking. However, this is not to way that the problems are solved by merely declaring an enemy to be a friend. New styles of formalism will be needed.

5 Examples of ongoing and evolving reasoning

In this section we briefly sketch several examples, illustrating the thesis that reasoning is necessarily an ongoing process, not only for reasons of computational limitations but because of the nature of the beast. The standard model is inadequate to properly represent any of these examples of commonsense reasoning; it will simply be unable to include the indicated inferences except in the presence of a contradiction, in which case because of omniscience it also sanctions *all* propositions as beliefs, thereby wiping out any useful distinctions on which to base recovery.

Language change It has been argued before [13, 19] that unlike the case of customary fixed formal languages, commonsense (or natural) language changes: new terms are coined or learned, old terms change meanings, etc. The reasoner must be able to reason about these changes, to incorporate them into her usage intelligently; and this involves noting tension (contradictions!) between usages. Noting that "John is tall" contradicts the personal

[5] As I myself have done; see [20]. Also see the introduction to [8].

observation that John is short, she starts to wonder whether these might be two different persons named "John" (see [15]).

Interpreting orders Your boss tells you (a personnel manager) never to hire high-school dropouts. One day a job candidate comes to your office. The interview goes fine and you note that he has a PhD. Then the next day you see that he in fact dropped out of high school, drifted for a few years, then managed to get a BS, MS and PhD with a fine scholastic and employment record. Do you hire him or not? Commonsense says this not what your boss meant by "HS dropout". But you are a little nervous because you realize that there is a clash of meanings, and you want to check it out with your boss.

Taking advice Advice taking [11] involves trusting what others say. But they may contradict what you believe, and you need to realize this even of you do trust them, so you can remove the contradicted beliefs. This is not necessarily straightforward, since it may take some reasoning to find out the contradictions.

Correcting misinformation You are given the combination to a lock, but when you try it, it does not work: either you forgot it or was told it wrong. So, you do not give up in despair: you try variations, such as reversing the numbers. But this too involves first noting a clash of beliefs, and remembering the wrong combination in order to vary it. Thus memory of old (untrusted) beliefs is important.

This and the previous examples may lead the reader to think that it is the interaction of our reasoner with other reasoners—i.e., a communication situation—that produces the need for recognition of contradictions. The combination lock problem can easily be refashioned solely in terms of a single reasoner; we leave this as an exercise and instead present below a different single-agent example.

Correcting perceptual errors You are walking in the woods and come across a log with an unusual growth of wildflowers along one edge. Later on you see it again and decide you have walked in a circle. But then you are not sure: is it the same log? The flowers look larger. You decide that it is not the same log and that you have not walked in a circle.

This and the other examples above are cases of change of belief, what in the standard model goes on in the TMS phase (or in the interface between CSR and TMS phases). But CSR is needed during this change, for it is precisely what the reasoner must rely upon to adjudicate between competing candidates for "axioms".

6 Conclusions

CSR then is in large part the ability to keep a cool head in the face of confusing data, and to undertake efforts to sort through the data, resort to trial and error if need be, and come to useful conclusions. Recognition of confusion, stop-gap remedies (cease trusting contradictands and closely-implicated data), and clarity-seeking by means of the rest of one's data, are central parts of an overall strategy. But detailed resolution of the confusion is highly domain-specific and thus must be undertaken on the basis of either previous expertise, expert supervision, or trial-and-error, while all the time making full use of the reasoning engine. Additions to the engine are done by the engine, not by a separate module while the engine is turned off or idling.

Thus self-adjusting logics of confusion seem to be the order of the day. What form such logics may eventually take is far from clear. I note that OSCAR [21, 24, 22, 23] as well as active (step) logics [4, 3, 5, 16, 9] are beginnings. It is clear that human commonsense reasoning involves many conflict-driven changes of belief, and that this is in need of being better understood for both cognitive and robotic purposes.

References

[1] A. Arruda. A survey of paraconsistent logic. In A. Arruda, R. Chuaqui, and N.C.A. da Costa, editors, *Mathematical Logic ni Latin America*, pages 1–41. North-Holland, 1980.

[2] J. Doyle. A truth maintenance system. *Artificial Intelligence*, 12(3):231–272, 1979.

[3] J. Elgot-Drapkin. *Step-logic: Reasoning Situated in Time*. PhD thesis, Department of Computer Science, University of Maryland, College Park, Maryland, 1988.

[4] J. Elgot-Drapkin. Step-logic and the three-wise-men problem. In *Proceedings of the 9th National Conference on Artificial Intelligence*, pages 412–417, 1991.

[5] J. Elgot-Drapkin and D. Perlis. Reasoning situated in time I: Basic concepts. *Journal of Experimental and Theoretical Artificial Intelligence*, 2(1):75–98, 1990.

[6] M. Ginsberg, editor. *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann, 1987.

[7] R. Guha and D. Lenat. Cyc: a midterm report. *AI Magazine*, 11(3):32–59, 1990.

[8] Jerry Hobbs and Robert Moore, editors. *Formal Theories of the Commonsense World*. Ablex, 1985.

[9] S. Kraus, M. Nirkhe, and D. Perlis. Planning and acting in deadline situations. Presented at the AAAI-90 Workshop on Planning in Complex Domains, 1990.

[10] J. Laird, A. Newell, and P. Rosenbloom. Soar: an architecture for general intelligence. *Artificial Intelligence*, 33:1–64, 1987.

[11] J. McCarthy. Programs with common sense. In *Proceedings of the Symposium on the Mechanization of Thought Processes*, Teddington, England, 1958. National Physical Laboratory.

[12] J. McCarthy. Circumscription: A form of non-monotonic reasoning. *Artificial Intelligence*, 13(1,2):27–39, 1980.

[13] J. McCarthy and V. Lifschitz. Commentary on McDermott. *Computational Intelligence*, 3(3):196–197, 1987.

[14] D. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13(1,2):41–72, 1980.

[15] M. Miller. *A view of one's past and other aspects of reasoned change in belief.* PhD thesis, Department of Computer Science, University of Maryland, College Park, Maryland, 1993.

[16] M. Miller and D. Perlis. Presentations and this and that: logic in action. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, Boulder, Colorado, 1993.

[17] M. Minsky. A framework for representing knowledge. In P. Winston, editor, *The Psychology of Computer Vision.* McGraw-Hill, 1975.

[18] N. J. Nilsson. Artificial intelligence prepares for 2001. *AI Magazine*, 4(4):7–14, 1983.

[19] D. Perlis. *Language, Computation, and Reality.* PhD thesis, Department of Computer Science, University of Rochester, Rochester, NY, 1981.

[20] D. Perlis. Languages with self reference I: Foundations. *Artificial Intelligence*, 25:301–322, 1985.

[21] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[22] John Pollock. *How to build a person.* MIT, 1989.

[23] John Pollock. Oscar: a general theory of rationality. *Journal of Experimental and Theoretical Artificial Intelligence*, 1(3):209–226, 1989.

[24] John Pollock. How to reason defeasibly. *Artificial Intelligence*, 57(1):1–42, 1992.

[25] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1,2):81–132, 1980.

[26] N. Roos. A logic for reasoning with inconsistent knowledge. *Artificial Intelligence*, 57:69–103, 1992.

[27] P. Rosenbloom, J. Laird, A. Newell, and R. McCarl. A preliminary analysis of the soar architecture as a basis for general intelligence. *Artificial Intelligence*, 47:289–325, 1991.

[28] S. Vere and T. Bickmore. A basic agent. *Computational Intelligence*, 6(1):41–60, 1990.