STEP-LOGICS: AN ALTERNATIVE APPROACH TO LIMITED REASONING

Jennifer Drapkin and Donald Perlis

University of Maryland
Department of Computer Science
College Park, Maryland 20742 USA
(301) 454-2002

Abstract: We propose that a new kind of logical study is appropriate to agents engaged in commonsense reasoning, namely, one that focuses on the steps of reasoning at any given time rather than the collection of all conclusions ever reached.

I. Introduction

The idea that commonsense reasoning is necessarily resource-bound, and in particular is not closed under ordinary logical consequence, has been suggested frequently. The kind of resource limitation that is most evident in commonsense reasoning (i.e., in reasoning about and within a real environment) is simply the very fact of passage of time while the reasoner reasons. There is not necessarily (or even likely) any fixed and final set of consequences that such a reasoning agent ends up with. In fact, the paradigm for such an agent would seem to be that suggested by Nilsson [1983], namely, a computer individual with a lifetime of its own. What is of interest for such an agent is not then its ''ultimate'' set of conclusions, but rather its changing set of conclusions over time. Indeed, there will be, in general, no ultimate or limiting set of conclusions.

The ''passage of time'' phenomenon is a limitation only in the following sense: the conclusions that may be logically (or otherwise) entailed by the agent's information (beliefs) take time to be derived, and time spent in such derivations is concurrent with changes in the world. Thus, it is not appropriate to spend hours figuring out a plan to save Nell from an onrushing train; she will no longer need saving by then (see [McDermott 1982] and [Haas 1985]). Even if the only changes are within the agent, this is still important, for it may be useful to know whether a problem is nearing solution, or if one has only begun initial explorations, and so on. That is, the agent should be able to reason about its ongoing reasoning efforts themselves. Thus, it is not so much an issue of weak resources as it is of a real-world fact about processes occurring over time. In fact, implemented reasoning systems obviously proceed to draw conclusions in steps; see for instance [Drapkin&Miller&Perlis 1986, Perlis 1981, Perlis 1984].

The literature contains a number of approaches to limited reasoning, apparently with these issues as motivational guides. However, most alternative approaches do not, in our view, carry this out to its logical (no pun intended) conclusions. With very little exception, the oversimplification of a ''final'' state of reasoning is maintained, and the limitation amounts to a reduced set of consequences rather than an ever-changing set of tentative conclusions. Thus, Konolige [1985a], for instance, studies agents with fairly arbitrary rules of inference, but assumes logical closure for the agents with respect to those rules, ignoring the effort involved in performing the deductions. Similarly, Levesque [1984] and Fagin&Halpern [1985] provide formal treatments of limited reasoning, so that, for instance, a contradiction may go unnoticed; but all the conclusions that *are* drawn are done so instantaneously, i.e., the steps of reasoning involved are not explicit. Lakemeyer [1986] deals with issues raised by then adding quantifiers, but does not address the issue that concerns us here. Vardi [1986] also deals with limitations on omniscience, but again without taking steps into account.

It is easy to provide examples in which the effort or time spent is crucial. The example of Nell is one illustration. For another, consider two agents, A and B, each of which has the intellectual ability (inference rules, etc.) to derive conclusion C. The two agents are told to try to determine whether C is true. But each derives that the other can derive C, and so each relaxes in the (mistaken) assumption that the other already must have derived C.

In part, this is a problem of modelling time, as has been studied by Allen [1984] and McDermott [1982]. However, there is more to it than this. Not only must the agent be able to reason about time, it must be able to reason *in* time. That is, as it makes more deductions, time passes, and this fact itself must be recognized. Otherwise we again face the prospect of losing Nell while deducing that it will take too long to get to a phone to call the train depot. We may even take too long to deduce that it will take too long! In other words, even the treatments of time in the literature are themselves still in the standard mold of unlimited or instantaneous reasoning.

II. Two languages for steps

This distinction leads to two directions for study. First, one would like an analytic formalism (AF) allowing us to determine what a given agent has and has not done at any given time. Second, the *agent* should also be able to reason (in some language/structure formalism RF) about what it has and has not done at any given time. These are obviously interrelated, and yet can be tackled somewhat independently. In particular, the ''analytic completeness'' of AF can be tackled without requiring the same of the agent, and *vice versa*. That is, we can seek a theory AF with the property that, for any given time i, and for any given wff α in the agent's language, AF should allow either a proof that the agent knows (has proved or otherwise determined) α at i, or a proof that the agent has not done so.

Here we propose some tentative characteristics of such languages as outlined above, including a precise characterization of analytic completeness. In [Drapkin&Perlis 1986] we have initiated a study with this aim in mind.

III. Step-Logics

We propose three major mechanisms to study as additions to RF: self (S), time (T), and retraction (R). Since it is important for the agent to reason about his own processes, a self, or belief, operator is needed. The agent would then have a way to talk (think) about his beliefs. In order for the agent to reason about time, a time operator is needed. Finally, since we want to be able to deal with commonsense reasoning, the agent will have to use default reasoning. That is, a particular fact is believed if there is no evidence to the contrary; however, later, in the face of new evidence, the former belief may be retracted. For this, we will need some kind of a retraction device. These new operators, of course, will necessitate corresponding changes in AF if the latter is to be able to ''keep up'' with a complete analysis of RF.

For the purposes of study, we have chosen to add these operators individually at first. We therefore propose a sequence of step-logics. We have come up with the following list. All the logics include time steps for AF. The list can be thought of as a progression or lattice in which later versions incorporate more features into RF. In SL_0, RF corresponds to a step-like propositional logic; later we illustrate this with a sample axiom.

SL_0

SL_1: S

SL_2: T

SL_3: R

SL_4: S, R

SL_5: S, T

SL_6: T, R

SL_7: S, T, R

It is again important to note the distinction between the *agent's* language and theory (RF), and *our* (the scientist's) language and theory (AF). The agent has one set of symbols, axioms, and rules, while we have another. For instance, "${}^{i}\alpha$" is used to indicate that the agent has proven α at time i. "α" is any wff in the *agent's* language. In SL_0, the "i" is just *our* notation; the agent has no way of knowing that it is at time i that he has proven α (he just "knows" α at time i). We, the scientists, however, can talk about what the agent has and has not proved through the use of |-, the first-order turnstile. For example, it might be the case that we have been able to prove that the agent has been unable to prove "P" in time i, where P is a predicate letter in the agent's language. (This in particular would be the case for an agent using ordinary propositional logic.) We would write this |- $\neg^{i}$P. We will continue the convention of using Greek letters for agent wffs. These also serve as terms of AF. To further distinguish AF and RF, we use "implies" and "not" as function symbols of AF to designate implication and negation, respectively, of the agent's wffs.

We have recently initiated the development of SL_0. SL_0 does not have S, T, nor R for the agent. To simplify it even more, it contains only propositions; it has no variables. As such, then, it is basically a formalism to help us, as scientists, to understand the reasoner. It does not allow the agent to do any reasoning about his own reasoning. We use the notation |- ${}^{i}\alpha$ to indicate that the agent has proven α in i steps. Note that α is a formula in the agent's language, but is treated as a constant in our language. We can think of "${}^{i}\alpha$" as an abbreviation for Thm(i,α), which refers to statements α that can be proven in the agent's theory in i steps. The goal is to design SL_0 to be strong enough so that for each i $\in$ **N**, and for each $\alpha \in$ L(RF) (where L(RF) is the agent's language), we have

$$SL_0 \vdash {}^i\alpha \quad \text{or} \quad SL_0 \vdash \neg {}^i\alpha.$$

This is our formal definition of ''analytic completeness''. See [Drapkin&Perlis 1986] for details of our current efforts to formalize such a theory SL_0. As a sample of the kind of axiom that we have found useful, we mention the following (similar to one found in [Haas 1985]):

$$\vdash (\forall i)(\forall j)\ [\ [{}^i\ \alpha\ \ \&\ {}^j\text{implies}(\alpha,\beta)\ \&\ i<k\ \&\ j<k] \rightarrow {}^k\beta]$$

Intuitively this expresses a version of the rule of *modus ponens*. Namely, if the agent ''knows'' (or has established) α at time i, and also knows at time j that α implies β, then it will know β at any time k greater than i and j.

IV. Conclusions

We have argued that for appropriate reasoning in the commonsense world, it is necessary to keep track of ones own steps of reasoning. Moreover, for us to be able to study such reasoners, it is necessary to have a formal description of ''step-reasoning'' so that it will be possible to determine whether, at a given moment, a given agent has (or has not) come to a certain conclusion. In the felicitous phrases of Doyle [1982] and Konolige [1985b], we need to formalize a ''rational'' or ''robot'' psychology. Here we have suggested particular avenues for doing just that, together with a ''completeness'' criterion.

Acknowledgements

This research has been supported in part by grants from the following organizations:
The IBM Corporation, for J. Drapkin
U. S. Army Research Office (DAAG29-85-K-0177), for J. Drapkin and D. Perlis
The Martin Marietta Corporation, for D. Perlis

Bibliography

(1) Allen, J. [1984] Towards a general theory of action and time. Artificial Intelligence, **23**, pp.123-154.

(2) Doyle, J. [1982] Some theories of reasoned assumptions: an essay in rational psychology. Dept. of Computer Science, CMU.

(3) Drapkin, J., Miller, M., and Perlis, D. [1986] A memory model for real-time commonsense reasoning. Draft.

(4) Drapkin, J. and Perlis, D. [1986] A preliminary excursion into step logics. *Proc. Intl. Symp. on Methodologies for Intell. Systems*, Knoxville, 1986.

(5) Fagin, R. and Halpern, J. [1985] Belief, awareness, and limited reasoning: preliminary report. IJCAI 85, pp.491-501.

(6) Haas, A. [1985] Possible events, actual events, and robots. Computational Intelligence, pp.59-70.

(7) Konolige, K. [1985a] A computational theory of belief introspection. IJCAI 85, pp.503-508.

(8) Konolige, K. [1985b] Experimental robot psychology. SRI International, Technical Note 363.

(9) Lakemeyer, G. [1986] Steps towards a first-order logic of explicit and implicit belief. *Proc. 1986 Conference on Theoretical Aspects of Reasoning about Knowledge*, pp.325-340.

(10) Levesque, H. [1984] A logic of implicit and explicit belief. Proceedings, 3rd National Conf. on Artificial Intelligence, pp.198-202.

(11) McDermott, D. [1982] A temporal logic for reasoning about processes and plans. Cognitive Science, **6**, pp.101-155.

(12) Nilsson, N. [1983] Artificial intelligence prepares for 2001. AI Magazine, **4**.

(13) Perlis, D. [1981] Language, computation, and reality. Ph.D. Thesis, University of Rochester.

(14) Perlis, D. [1984] Non-monotonicity and real-time reasoning. Proceedings, AAAI Workshop on Non-Monotonic Reasoning, New Paltz, Oct. 17-19.

(15) Vardi, M. [1986] On epistemic logic and logical omniscience, *Proc. 1986 Conference on Theoretical Aspects of Reasoning about Knowledge*, pp.293-305.