erase

A PRELIMINARY EXCURSION INTO STEP-LOGICS

Jennifer Drapkin and Donald Perlis

University of Maryland
Department of Computer Science
College Park, Maryland 20742

ABSTRACT: We have suggested that a new kind of logical study that focuses on individual deductive steps is appropriate to agents that must do commonsense reasoning. In order to adequately study such reasoners, a formal description of such ‘‘steps’’ is necessary. Here we carry further this program for the propositional case. In particular we give a result on completeness for reasoning about agents.

KEYWORDS: commonsense, time, reasoning steps

I. Introduction

In [4] we proposed that a new kind of logical study was appropriate to agents engaged in commonsense reasoning, namely, one that focuses on the steps of reasoning at any given time, rather than dealing with the collection of all conclusions ever reached. For this, one would like an analytic formalism (AF) allowing us to determine what a given reasoner has and has not done at any given time. Also, the **reasoner** should also be able to reason (in some language/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 reasoning 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 pursue the study of AF in the case of a propositional reasoner, i.e., an agent with propositional logic as its inferential mechanism.

We contend that the kind of resource limitation that is most evident in commonsense reasoning is the passage of time while the reasoner reasons. There is not necessarily any fixed and final set of consequences with which such a reasoning agent ends up. In fact, the paradigm for such an agent would seem to be that suggested by Nilsson [15], 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, in general, be no ultimate or limiting set of conclusions.

The above phenomenon is a limitation in the following sense: the conclusions that may be logically 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. Even if the only changes are **within** the agent, still this is 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.

The current paper is an extension of [4]. In that paper we began some formal details of AF. In fact, we proposed a list (actually a lattice) of step-logics, arranged in increasing sophistication, from SL_0 to

SL_7. The former has simply a propositional reasoner as agent, with no internal knowledge of steps; only the analytic formalism (AF) records the agent's steps. The latter not only allows the agent first-order reasoning, but also knowledge of self and time, and the ability to retract beliefs. A key feature of step-logic is that it lends itself naturally to belief revision, in that the agent is never faced with the totality of all logical consequences of its beliefs at any moment. Rather, the belief set is finite (but changing); this allows the possibility of a genuinely computational solution to introspection, and decision as to what beliefs (if any) are to be removed. We will not elaborate further on this here since in this paper we will focus solely on SL_0.

The remainder of our treatment is as follows: In section II we present further motivation and related approaches in the literature. Section III contains some formal directions we feel are worth exploring. In section IV we present some results of having pursued one such direction, including the aforementioned analytic completeness.

II. Motivation

A puzzle that a reasonable model of knowledge ought to be able to solve, is the **three-wise-men** problem. We present a variation of this classic problem which was first introduced by McCarthy [11]. A king wishes to know whether his three advisors are as wise as they claim to be. Three chairs are lined up, all facing the same direction, with one behind the other. The wise men are instructed to sit down. The wise man in the back (A) can see the backs of the other two men. The man in the middle (B) can only see the one wise man in front of him (C); and the wise man in front (C) can see neither A nor B. The king begins with five cards, three white, and two black. He throws two (of unknown color) away, then places one card, face up, behind each of the three wise men. The men are given 30 minutes to determine the color of the card that sits behind his own chair. The room is silent; then, after 29 minutes, C says 'My card is white!'.

The reasoning that supposedly occurred is as follows. Because the king started out with three white and two black cards, then threw two away, each wise man must realize there is at least one white card. If the cards of B and C were black, then A would have been able to announce immediately that his card was white. They all realize this (they are all truly wise). Since A kept silent, either B's card is white, or C's is. At this point B would be able to predict, if C's were black, that his card was white. They all realize this. Since B also remains silent, C knows his card must be white.

One of our eventual goals is to develop a formalism which is capable of performing the above reasoning. We do not achieve this here. Our current results contribute to the analytic half of the problem. We have been able to achieve analytic completeness for propositional agents; that is, for any agent wff α and any time i, our formalism (SL_0) can either prove ${}^i\alpha$ or can prove its negation. This means that the agent's reasoning over time is completely characterized, in the sense that we know what it has and has not established at each moment.

We quickly mention further examples in which the effort or time spent is crucial. 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 Haas [7] and McDermott [12]). For another example, consider two agents, A and B, each of which has the intellectual ability (inference rules, etc.) to derive conclusion C, and to reason about the other's reasoning abilities. 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, these are problems of modelling time, as has been studied by Allen [1] and McDermott [12]. 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 reasoning.

The literature contains a number of approaches to limited reasoning, apparently with these issues as motivational guides. However, 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 [8], 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 [10] and Fagin and Halpern [6] provide formal treatments of limited reasoning, so that, for instance, a contradiction may go unnoticed; but the conclusions that **are** drawn are done so instantaneously, i.e., the steps of reasoning involved are not explicit. The logics of Levesque, and Fagin and Halpern deal with propositional reasoners. Lakemeyer [9] deals with issues raised by then adding quantifiers, but does not address the issue that concerns us here. Vardi [16] also deals with limitations on omniscience, but again without taking steps into account.

III. Step-Logics

As was indicated in section I, we would like to build self (S), time (T), and retraction (R) operators into RF. These of course will necessitate corresponding changes in AF, in order for AF to be able to ‘‘keep up’’ with a complete analysis of RF. The agent (using RF) would then be able to talk about his beliefs, reason about time, and retract former beliefs. For the purposes of study, however, it is useful to add these operators individually at first. We have therefore suggested a lattice of step-logics as mentioned earlier.

We note again the difference between the **agent’s** language/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 by us to indicate that the agent has proven α at time i; ‘‘α’’ is any wff in the **agent’s** language. 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 propositional letter in the agent’s language. We would write this |- $\neg^i$P. We will continue the convention of using Greek letters for agent wffs. To further differentiate AF and RF, we use ‘‘implies’’ and ‘‘not’’ as function symbols of AF to designate implication and negation, respectively, of the agent’s wffs. L(RF) is used to indicate the set of agent wffs.

It is worth making a distinction here between **traditional** meta-theorems and the meta-theorems in which **we** are interested. Those theorems proved about conventional logics are asymptotic in nature, that is, they demonstrate useful properties of the **limiting** case. The meta-theorems we wish to prove about SL_0 and all subsequent step-logics, on the other hand, concern the **bounded** case. For the most part, we are not interested in what the agent may eventually know, but rather what it may know within a finite amount of time, whether that be 10 seconds, or 10 years.

One hope of ours with each of the step-logics is to be able to characterize the agent by determining what it does or doesn’t know at any given time. That is, for an arbitrary agent wff α, and for any time i, we would like to say |- $^i\alpha$ or |- $\neg^i\alpha$, in any of the logics. Such a result, we call ‘‘analytic completeness’’: the logic can completely analyze the agent’s reasoning behavior.

We are not concerned, however, with showing the following asymptotic result,

$$|\!-\ (\exists i)^i\alpha \quad \text{or} \quad |\!-\ \neg(\exists i)^i\alpha.$$

Although this is decidable in the propositional case (tautology testing will do), this, in fact, is undecidable in predicate calculus. However, we are only interested in what the agent has been able to prove in a finite amount of time, and this **is** decidable (that is, analytic completeness holds), at least for the particular version of a propositional agent that we have chosen to study.

Certain asymptotic results, however, are sometimes useful and/or interesting. For example, Theorem 2 stated below is a theorem dealing with the limiting case. It says that all agent conclusions are tautologies, and was used to arrive at the corollary which states that a single propositional letter is never proven. The contra-positive of Theorem 2 would be interesting to verify as well, namely, that all tautologies are eventually proven.

We are focussing on the AF in this paper, as an initial aspect of a larger study. Development of the RF in a way suitable for the kind of problem given in section II will be presented in future work. However, analytic completeness, which is a property of the AF, is important to the entire undertaking in that it establishes that we have indeed represented an agent. That is, this guarantees that an agent has been adequately and fully specified, short of actually building/simulating its behavior.

IV. SL_0

In SL_0 the agent has neither S, T, nor R. To simplify it even more, the agent uses only propositional logic. (For definiteness, we have arbitrarily picked one of the standard axiomatizations in the literature.) As such, then, SL_0 is basically a formalism to help us to understand the reasoner. It does not allow the agent to do any reasoning about his own reasoning. 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. We wish SL_0 to be powerful enough so that for each i ∈ N (the natural numbers), and for each $\alpha \in$ L(RF),

$$SL_0 \mid\!\text{-}\ {}^i\alpha \quad \text{or} \quad SL_0 \mid\!\text{-}\ \neg{}^i\alpha.$$

As mentioned, we call this analytic completeness. In this section we sketch a version of SL_0 adequate for demonstrating this result. As might be expected, the hard part is determining when a wff of L(RF) is **not** deduced after i steps.

We first sketch intuitively the operation of our intended agent. For each wff α in the agent language L(RF), and for each time (step) i ∈ **N**, $^i\alpha$ is to hold (i.e., the agent has deduced α by time i) if and only if there is a formal proof of α in i or less steps using only **modus ponens** and whatever axioms can be retrieved in i steps. Here axioms are conceived to be constructed by the agent little by little. This has some similarity to Fagin-Halpern's notion of awareness. More particularly, the agent begins with no axioms at all, and at any step i has access to all old conclusions as well as any axioms using no more than the first i proposition letters and no more than i connectives (instances of ¬ and →). In order to distinguish the agent's wffs from our own (i.e., wffs of AF or SL_0), we write not(α) for the agent's negation of α, and implies(α,β) for the agent's representation of α implies β. In SL_0, then, **not** and **implies** become function symbols. Also, for agent wffs to appear as constant terms in SL_0, proposition letters must be constant symbols of L(AF).

SL_0 is given enough arithmetic to reason about steps as integers. In addition, the following principal axioms and axiom schemata are given:

AX: MY$(\forall\alpha)$ [Ax(α) ↔
$\exists\beta)(\exists\gamma)(\alpha=\text{implies}(\beta,\text{implies}(\gamma,\beta)))$
$\exists\beta)(\exists\gamma)(\alpha=\text{implies}(\text{implies}(\text{not}(\gamma),\text{not}(\beta)),\text{implies}(\text{implies}(\text{not}(\gamma),\beta),\gamma)))$
$\exists\beta)(\exists\gamma)(\exists\delta)(\alpha=\text{implies}(\text{implies}(\beta,\text{implies}(\gamma,\delta)),\text{implies}(\text{implies}(\beta,\gamma),\text{implies}(\beta,\delta))))$]

AX says that axioms are of the three tautology types as in [13].

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

*This is a version of **modus ponens.***

THM: $(\forall\alpha)(\forall i)[{}^{i}\alpha\ \ \leftrightarrow [\ \mathrm{MZ}[\mathrm{Ax}(\alpha)\ \&\ (l(\alpha) \leq i)\ \&\ (p(\alpha) < i)\ ]$
$(\exists j)(\exists k)(\exists\beta)({}^{j}\beta\ \&\ {}^{k}\mathrm{implies}(\beta,\alpha)\ \&\ j<i\ \&\ k<i)],$

*THM defines what it means for α to be proven by the agent at time i: either α is an axiom, and has been ‘‘fed in’’, or α has been derived through **modus ponens** from previous steps.*

TabulaRasa: $(\forall\alpha)(\forall i)[{}^{i}\alpha \rightarrow (i >= 0)]$.

*The agent knows nothing before time step 0.

LN1: $(\forall\alpha)(\mathrm{PL}(\alpha) \leftrightarrow l(\alpha) = 0)$.

LN2: $(\forall\alpha)[l(\mathrm{not}(\alpha)) = l(\alpha) + 1]$

LN3: $(\forall\alpha)(\forall\beta)[l(\mathrm{implies}(\alpha,\beta)) = l(\alpha) + l(\beta) + 1]$

LN4: $(\forall\alpha)(l(\alpha) >= 0)$

The function $l(\alpha)$ returns the length of α, i.e. the number of connectives in α.

P1: $(\forall\alpha)[(\mathrm{PL}(\alpha)\ \&\ p(\alpha)=i) \leftrightarrow (\alpha) = P_i)]$

P2: $(\forall\alpha)[p(\mathrm{not}(\alpha)) = p(\alpha)]$

P3: $(\forall\alpha)(\forall\beta)[p(\mathrm{implies}(\alpha,\beta)) = \max(p(\alpha),p(\beta))]$

P4: $(\forall\alpha)[p(\alpha) >= 0]$

The function $p(\alpha)$ returns the maximum index of all the propositional letters in α.

MAX1: $(\forall i)(\forall j)[i>=j \leftrightarrow \max(i,j)=i]$

MAX2: $(\forall i)(\forall j)[\max(i,j)=\max(j,i)]$

The function max returns the maximum of its two arguments.

EQ1: $s \neq t$, for all distinct propositional letters $s,t \in L(RF)$

EQ2: $s \neq \mathrm{implies}(t,u)$, if s is a propositional letter $\in L(RF)$, and $t,u \in L(RF)$.

EQ3: $s \neq \mathrm{not}(t)$, if s is a propositional letter $\in L(RF)$, and $t \in L(RF)$.

EQ4: $\mathrm{implies}(s,t) \neq \mathrm{not}(u)$, for all $s,t,u \in L(RF)$.

EQ5: $\mathrm{implies}(s,t) = \mathrm{implies}(u,v) \leftrightarrow s=u\ \&\ t=v$, for all $s,t,u,v \in L(RF)$.

EQ6: $\mathrm{not}(s) = \mathrm{not}(t) \leftrightarrow s=t$, for all $s,t \in L(RF)$.

EQ1, EQ2, and EQ3 say that distinct agent wffs represent distinct objects. EQ4 says that a wff whose last connective is an implication arrow cannot be equal to a wff whose last connective is a not symbol. EQ5 (EQ6) says that in order for two wffs to be equal, where the implication arrow (not symbol) is the last connective that was used to join each of the wffs, their arguments must be equal.

TFCN: $(\forall x)[\ \mathrm{Tfcn}(x) \rightarrow .\ \mathrm{MW}(\forall\alpha)(\forall\beta)[\mathrm{True}(x,\mathrm{implies}(\alpha,\beta))\ \leftrightarrow .\ \mathrm{True}(x,\beta)\ \mathbf{v}\ \neg\mathrm{True}(x,\alpha)]$
$(\forall\alpha)[\mathrm{True}(x,\mathrm{not}(\alpha)) \leftrightarrow \neg\mathrm{True}(x,\alpha)]\ ]$

TFCN says that truth-functions behave properly with respect to connectives.

TAUT1: $(\forall\alpha)[\ \text{Taut}(\alpha) \leftrightarrow (\forall x)(\text{Tfcn}(x) \rightarrow \text{True}(x,\alpha))\ ]$

TAUT2: $(\forall\alpha)[\text{Ax}(\alpha) \rightarrow \text{Taut}(\alpha)]$

TAUT1 and TAUT2 say, respectively, that tautologies are wffs that are true under all truth-functions, and that axioms are tautologies.

PL1: $(\forall\alpha)[\text{PL}(\alpha) \rightarrow (\exists x)[\text{Tfcn}(x)\ \&\ \neg\text{True}(x,\alpha)]\ ]$

PL2: $\text{PL}(P_i)$, for all $i \in \mathbf{N}$.

*PL1 says that for each propositional letter there is a truth-function making it false. PL2 says that P_0, P_1, P_2, ... are propositional letters. Note that this does not give **all** the usual truth-functions, but it is sufficient for our purposes.*

We then have the following results:

*Theorem*1: $SL_0 \vdash (\forall\alpha)(\forall\beta)[\ [\text{Taut}(\alpha)\ \&\ \text{Taut}(\text{implies}(\alpha,\beta))\ ] \rightarrow \text{Taut}(\beta)\ ]$.

*Theorem*2: $SL_0 \vdash (\exists i)\,{}^{i}\alpha \rightarrow \text{Taut}(\alpha)$, for any wff $\alpha \in \text{L(RF)}$.

The only wffs an agent will ever prove are those that are tautologies.

Lemma: $SL_0 \vdash \neg\text{Taut}(P)$, for any propositional letter $P \in \text{L(RF)}$.

P is not a tautology, for all propositional letters P.

Corollary: $SL_0 \vdash (\forall i)\neg\,{}^{i}P$, for any propositional letter $P \in \text{L(RF)}$.

The agent will never be able to prove P, where P is a propositional letter. This follows immediately from Theorem 2 and the lemma.

Monotonicity Lemma: $SL_0 \vdash {}^{i}\alpha \rightarrow {}^{(i+1)}\alpha$

Once a belief is held, it remains.

Access Lemma: $SL_0 \vdash {}^{i}\alpha \rightarrow [\text{l}(\alpha) \le i]\ \&\ [\text{p}(\alpha) < i]$.

Those theorems proved by the agent at time i are only those with length less than or equal to i, and containing no propositional letters with index greter than i. This follows immediately from THM.

Lemma: If $SL_0 \nvdash \text{Ax}(\alpha)$ then $SL_0 \vdash \neg\text{Ax}(\alpha)$.

SL_0 can always prove whether or not a wff is an axiom (as defined by Ax).

Boundedness Lemma: Let $i \in \mathbf{N}$, $i >= 2$. Then $\exists\alpha_1 ... \alpha_{n_i} \in \text{L(RF)}$, such that

$$SL_0 \vdash (\forall\alpha)[\ {}^{i}\alpha \leftrightarrow [\ \alpha=\alpha_1\ \mathbf{v}\ ...\mathbf{v}\ \ \alpha=\alpha_{n_i}\ ]\ ].$$

We can then prove the following result.

Analytic Completeness Theorem: For each $i \in \mathbf{N}$, and for each $\alpha \in \text{L(RF)}$,

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

SL_0 can characterize exactly what has and has not been proved at any given time i.

Proofs of the above results tend to be rather long, and proceed mainly by induction on i and/or the number of connectives in α.

V. Conclusions

We have argued that in order to do appropriate reasoning in the commonsense world, it is necessary to keep track of one's own steps of reasoning. Moreover, for us to be able to study such reasoners, it is necessary to have a formal description. Here we have suggested particular avenues for doing just that. We have developed the first in a sequence of formalisms that allows us to keep track of the agent's steps of reasoning. The analytic completeness criterion was obtained for this logic.

This research has been supported in part by the following institutions:

The U.S. Army Research Office (DAAG29-85-K-0177)

The Martin Marietta Corporation.

The IBM Corporation.

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. Technical report, Univ. of Maryland.

(4) Drapkin, J. and Perlis, D. [1986] Step-logics: an alternative approach to limited reasoning. **Proc. European Conf. on Artif. Intell.**, 1986.

(5) Drapkin, J. and Perlis, D. [1986] Analytic completeness in SL_0. Technical report, Univ. of Maryland.

(6) Fagin, R. and Halpern, J. [1985] Belief, awareness, and limited reasoning: preliminary report. **Proc. 9th Int'l Joint Conf. on Artificial Intelligence**, 1985, pp.491-501.

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

(8) Konolige, K. [1985] A computational theory of belief introspection. **Proc. 9th Int'l Joint Conf. on Artificial Intelligence**, 1985, pp.503-508.

(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. **Proc. 3rd National Conf. on Artificial Intelligence**, pp.198-202.

(11) McCarthy, J. [1978] Formalization of two puzzles involving knowledge. Unpublished note, Stanford University, Stanford, California.

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

(13) Mendelson, E. [1972] **Introduction to mathematical logic**, van Nostrand.

(14) Moore, R. [1985] A formal theory of knowledge and action. **Formal Theories of the Commonsense World**, Ablex Publishing Company, pp. 319-358.

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

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