Reprinted from International Journal of Intelligent Systems 6(1991)443--456

Stop the World---I Want to Think*

Donald Perlis†
Department of Computer Science
and
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742

Jennifer J. Elgot-Drapkin
Department of Computer Science and Engineering
College of Engineering and Applied Sciences
Arizona State University
Tempe, AZ 85287-5406

Michael Miller‡
Department of Computer Science
University of Maryland
College Park, MD 20742

*With apologies to Leslie Bricusse and Anthony Newley

†Supported in part by ARO research contract no. DAAL03-88-K0087.
‡Supported in part by ARO research contract no. DAAL03-88-K0087.

Abstract

Reason-based actions plunge the reasoner into temporal considerations from all angles. We see this not only when time enters explicitly into the problem statement, but also in formal robot blocks-world scenarios, in the Yale Shooting Problem and other associated versions of the frame problem (e.g., [Hanks and McDermott, 1986]), in various specialized actions (e.g., hiding, as in [Allen, 1984]), and so on. In short, where there is action, there is time, and where there is time, there is a potential need for reasoning about time.

Where, then, is the action? It certainly includes the usual overt physical acts of motion, and also certain covert behaviors such as hiding or watching. In these of course time is important.

But there is another angle that is not usually noted, one that we have been exploring for the past several years ([Drapkin and Perlis, 1986b],[Drapkin and Perlis, 1986a], [Elgot-Drapkin, 1988],[Elgot-Drapkin and Perlis, 1990]). Namely, action also occurs in the form of mere thinking or reasoning. Moreover, the very same temporal considerations apply to this reasoning behavior. This leads us to view reasoning itself as a kind of action, with the obvious yet non-trivial consequence that our reasoning goes on ''as the world turns''. The present paper offers various arguments in support of this position.

1 Introduction

Actions in AI are traditionally viewed as separate from the planning process that leads to those actions. Even when the two are intertwined, as in real-time, dynamic, or reactive planning, still the planning effort is usually treated as a different kind of beast, and not (entirely) as an action itself.[1] We argue that this is a mistake. Recognizing oneself as a reasoner, where one is engaged in an *activity* of reasoning (as well as other actions), is crucial to many tasks. Planning often requires planning to plan (i.e., to reason) about the task at hand.

In [Drapkin and Perlis, 1986b] our interest in this theme is enunciated. Some preliminary results are reported in [Drapkin and Perlis, 1986a]. More recently, [Elgot-Drapkin, 1988] gives a wider variety of technical and implementational results, many of which are published in [Elgot-Drapkin and Perlis, 1990]. The aim of the current paper is to step back and re-assess the conceptual basis for this work to see how it fits into broader themes and research directions.

While the format of our presentation may seem unusual, we feel that it suits the purpose of this paper well. In particular, we will present many short sections that provide arguments in favor of our position or answers to rhetorical questions, leading eventually to our conclusions

[1] For instance, [Dean and Boddy, 1988] (page 50) and [Russell and Wefald, 1989] (page 402) both decide to ignore at least certain aspects of the time taken to plan.

concerning formal and implementational approaches and our experiences so far in this endeavor. We do *not* attempt to present details of our formalisms, which are already in several places in the literature. (See [Drapkin and Perlis, 1986b], [Drapkin and Perlis, 1986a], [Elgot-Drapkin, 1988],[Elgot-Drapkin and Perlis, 1990].) Rather, our goal is to provide a *philosophical* basis for this work.

The reader should be aware that we have already developed a full formal theory that obeys the strictures urged here, and implemented it and applied it to various problems in commonsense reasoning.[2] The papers in which these technical results are presented do not, however, go very deeply into the motivations behind the formal details. Filling this gap is the purpose of the present work. Prior familiarity with the formal details is not necessary for understanding the present paper, which we hope will make serious points independently of our formal work---points which we think have very broad significance beyond our own formalisms.

In short, we think that time must be incorporated into much more work in AI, not only in its commoner form of temporal reasoning (reasoning about time) but also in the form we urge here: reasoning about time *in* time. This point is the main one we wish to argue.

2 Time and temporal reasoning

Temporal reasoning, as it is usually understood, has to do with reasoning about time, in a timeless reasoning present. Some important work in this area includes [McDermott, 1982] and [Allen, 1984]. What we call reasoning in time is a rather different matter: it merely reflects the commonplace truism that reasoning itself goes on in time, no matter what the reasoning is about. Thus we have the two, superficially unrelated, topics of reasoning about time and reasoning in time.

But now we introduce *reasoning in time about one's own reasoning in time*. This is what we claim is needed for realistic real-time systems. If a reasoner is midway through a problem, she may pause to assess where she is, and as she does, several seconds (or minutes) may pass. She must take account of that, and decide to push ahead rather than spend more time assessing.

Is this just dovetailing object and meta levels? No, it is not, for the assessing is what takes account of its *own* squandering of temporal resources; there is no retreat to yet another meta-meta

[2]For instance, Moore's *Brother Problem* and the *Three-wise-men Problem*.

level for that. (If there were, we'd be in a worse mess, even further from the real-time system we wish to achieve.)

3 Actions

What is an action? We offer the following (admittedly loose) definition:

> An *action* is that which takes time---or goes on in time---and is at least partly under one's control, i.e., the object of one's planning.

This includes waiting, hiding, but perhaps not sleeping (at least not while it is being done!). It also includes all manner of common behaviors, such as running, eating, pushing, sitting, driving. One might even argue that it includes such states as hating or wanting, although this is less clear, since it is not obvious that these are under one's control (can one plan to hate?). In any case, our definition seems to include most physical behaviors, that is, things with environmental impact. Let us call such actions *base actions*.

Most research into actions (and planning, which is the process of choosing actions) has treated only base actions. Indeed the reasoner typically regards planning (i.e., thinking) as a separate sort of endeavor. (To call it an activity at this point would be to beg the question, but that is what it is, as we shall argue next!)

4 Reasoning is action

Now that we have a notion of action, how can we hope to encompass reason within such a notion? Planning (i.e., reasoning) is not a base action. But it is really a kind of action by our definition: it goes on in time, and we are (at least some of the time) in control of it. A definitional argument has no impact, however, unless we can show it to be useful.

How can we see this? What is it about actions that make them the objects of planning? The key idea implicit in our definition is that we need to understand certain features of actions if we are to make an intelligent plan (i.e., choice of actions). But the very same can be said of planning itself! We need to understand certain features of planning if we are to make an intelligent choice in going about our planning. That is, we often need to plan to plan (or reason). Thus reasoning (at

least reasoning that formulates plans) is an action, not only by our definition, but by the notion of utility behind it.

5 Deadlines

Reasoning about plans often involves deadlines. Not only must base actions be produced on a schedule, but planning as well is subject to a time constraint. If we take too long to plan, the time for acting (i.e., for the base action) may have passed. The *Nell & Dudley Problem* described in Section 14 provides a clear illustration of this.

6 Infinite regress

If we need to plan to plan, then don't we also need to plan to plan to plan, etc? And if so, won't this undermine the whole point of our real-time outlook? That is, won't the reasoner be led into so much meta-planning that all deadlines will be missed?

Clearly we must cut off the planning somewhere, to be sure. But this can be determined relative to the needs of the problem at hand. We don't always need to plan to plan. Even less frequently do we need to plan to plan to plan. In those cases where we really do need to plan to plan to plan, we had better not have a tight deadline, or we will likely fail to meet it.

7 Planning to reason

A simple example of planning to reason is the following. I plan to go to the store to compare costs and then decide what to buy. Here I see myself as a reasoner, as an agent with the ability to carry out a task of reasoning (perhaps as a subgoal of a larger task). At some point in my later enactment of this plan, I enact the part of comparing and deciding, clearly themselves reasoning processes. Now, this may be easy to admit, without granting it any importance for understanding the principles behind planning and acting. So we must argue more.

8 As the World Turns

The usual treatments of temporal reasoning assume a kind of stasis, as if the world stops while the reasoner reasons. This supposedly allows all (infinitely many) valid conclusions about the future to be reached, and an appropriate plan for a course of action to be selected, without the world having changed from under the reasoner while reasoning was taking place. But clearly time does not stand still while planning goes on. The world changes, if only by virtue of time itself being different as we finish our episode of thought from what it was when we began. Instead of a *stop-the-world* world, we live in an *as-the-world-turns* world.

9 Litterbugs: an issue of space

Many of the same arguments we give against the *stop-the-world* approach can also be levied against the *litterbug* approach, i.e., the assumption of unlimited space for storing beliefs, etc. We are very guilty of being litterbugs in our own formalism (‘‘step-logics’’, which we will describe briefly toward the end of this paper): although our logics have at any given moment only a finite set of beliefs (in principle as well as in implementation), still finite can be very large! We are perhaps less guilty than the stop-the-worlders with their genuine infinitudes. Nevertheless no version of unbounded litterbuggery is consistent with our underlying philosophy.[3]

10 Just an implementation issue?

Still a quiet voice murmurs to us, isn’t this really an implementation issue? Aren’t real-time concerns a matter for engineers to worry about, not theoreticians? Isn’t the same true for space (litterbug) concerns? Can we really mix theory and practice together and have something coherent? Can time considerations really be fruitfully allowed into the study of reasoning *per se*, without thereby undoing all hope of having something principled, theoretical, foundational? Can’t we assume faster processors will allow thinking (planning) to occur, in effect, instantaneously?

No, because, firstly, some problems are undecidable. The reasoner will *never* finish; she will never reach all valid conclusions.[4]

[3]While we have ideas on how to deal with this, we have ignored the issue in the interests of focusing our energies on one theme at a time (pun intended!).

[4]This holds even for merely semi-undecidable problems.

Secondly, it is very doubtful that future progress in processor speed, or in parallel architectures, will allow arbitrarily fast planning.[5] Thus if we are ever to build intelligent machines, or understand ourselves as intelligent information processors, it will have to be, in part, in terms of reasoning going on in time.

Thirdly, regarding human cognitive modelling, it is clear that *we* make regular use of time-considerations as we reason. Consider our earlier example of comparing prices: if we need to purchase an item very soon, we may plan to do only a little comparing before choosing, rather than first thinking of all possible stores we could check. See Section 14 for other examples.[6]

11 Application to default reasoning

The usual formal treatments take, in effect, the *stop-the-world* approach, in that they assume that all conclusions theoretically available to the reasoner are in fact achieved, and that the world has not changed in the meantime. This problem (and it is a problem, as we argued in Section 8) disappears when we cease demanding inferential omniscience (another name for the *stop-the-world* approach). That is, an assumption of inferential omniscience amounts to assuming the reasoner has enough *time* to work out all consequences (leading to closure under the same) of her beliefs, which for any interesting set of inference rules will be very large and usually infinite. But to perform an infinite or even very large number of inferences will take too long in general for effective reasoning coupled with subsequent actions. Hence, implicitly the reasoning is done while the world is (assumed to be) stopped.

Consider what happens in a typical default setting. We want to conclude that, since we know nothing unusual about entity Q with respect to some property P, then we may as well assume Q is typical with respect to P. But if we are omniscient, when we proceed to determine what we do and don't know, we will have to decide all that we could possibly come to know given all the time in the world---in general an infinite (and undecidable) set.

[5]We could cite quantum limits here---e.g., see [Simon, 1978]---but that seems like overkill. Does anyone really believe that computation even in principle can be arbitrarily fast and yet limited to a modest-sized device such as a mobile robot or living brain?

[6]However, the issue of principled theories is a legitimate concern. We ourselves were concerned with this as we started our investigation: how can we formulate a logic in which certain conclusions (e.g., what time it is) change out from under us as reasoning proceeds? To our pleasant surprise, our formulation has turned out very nicely. It is theoretical and yet implementation-situated at the same time. We give a brief sketch of our formalism in Section 15.

On the other hand, if we are *not* omniscient, then we can reformulate the default idea as follows: Since we have not been able to think of anything unusual *so far* about Q with respect to P, then we may as well assume Q is typical with respect to P. Now there is nothing at all that needs to be computed, other than an assessment of what we have already stored away from past computation (reason). Since at any given moment this will be a *finite* set of conclusions from past computation, our task is computable.

12 Correcting default conclusions

In the rapidly expanding literature on default (or nonmonotonic) reasoning, little attention has been paid to the problem of what to do once a default conclusion has been found to be in error. Indeed, given the omniscient approach this problem is truly severe, since an infinitude of conclusions may have to be undone.

Traditional implementation-oriented, yet logic-based, approaches tend to treat this ''error recovery'' problem as a separate routine from the reasoning that produced the error in the first place. (See, for example, [Doyle, 1979] and [deKleer, 1986].) But in general the very same commonsense knowledge used to draw a default conclusion will also be needed to figure out what to believe when the default is found to be in error. Thus recovering from error should not in principle require separate treatment.[7]

Consider now the picture if we view reasoning as going on in time, with an always finite[8] set of conclusions at any moment. If at some point we either are presented with new data---or through reason or recall come upon further inferences---showing a prior conclusion about Q being typical with respect to P to be false, we need not undo (or inspect) an infinitude of conclusions, nor need we even switch into another mode (of error-correction). Rather, we can refrain from believing the false conclusion from now on, but still retain as a remembered belief that we had formerly believed it.[9]

[7]Other implementation-oriented approaches, such as SOAR ([Laird *et al.*, 1987]) and our own ([Elgot-Drapkin *et al.*, 1987]), avoid all the problems we are discussing. In the process, however, they are no longer ''theories'', in the sense of embodying concise, general, and powerful explanatory principles as to the nature of reasoning. Is there then a happy medium? We claim there is.

[8]This is of course a necessary condition for any real-world implementation, if we regard conclusions as physically stored entities; and it also is obeyed by our own formalism.

[9]This sort of analysis is hinted at for formal theoretical reasons in [Perlis, 1986] and [Perlis, 1988].

One way to think of this is to imagine new data in conflict with already existing data, providing a momentary contradiction. This can be dealt with, roughly speaking, by looking for direct clashes between any two elements in the current set of beliefs (conclusions) at every ‘‘step’’ in the reasoning. If found, the two beliefs in question are not used to draw further inferences as long as they both are present. If one of the two maintains itself longer than the other (supposedly due to more robust evidence), then eventually only it remains, and now can be used to draw further conclusions. This has in fact been carried out formally; details can be found in [Elgot-Drapkin, 1988]. This leads us to the discussion of inconsistencies.

13 Application to inconsistency

A traditional scare-tactic in the literature is to suggest that someone's formal theoretical apparatus for commonsense reasoning contains inconsistencies under certain conditions. This should not be regarded as so unsettling; to the contrary, the cropping up of inconsistencies seems almost the hallmark of commonsense reasoning. Of course, we humans generally know what to do about inconsistencies: we note their existence and stay clear, at least until a way to dissolve them comes to mind. This is exactly what we have aimed at in the method sketched in the previous section.

Why is this so alarming in more traditional approaches? This again has to do with omniscience. If literally *all* valid conclusions emanating from one's beliefs are already known (i.e., already among one's beliefs), and if there is an inconsistency there, then one's belief set contains every single assertion whatsoever---in particular, every assertion *and* its negation! The set is swamped with utter inconsistency, not just a few isolated direct contradictions.[10]

On the other hand, the time-tempered gradual approach we urge does not have the time to carry out all these absurd (even if valid) deductions. The production of inferences is not separated at all from the inspection of possible contradictands, nor from the refusal to base further inferences on such contradictands.

As for *how* to deal with contradictands when they arise, we suspect that there are as many issues (and methods) here as there are situations in which contradictions can arise. We have explored some approaches relevant to particular commonsense problems (see below). However, we are not suggesting that any one approach will be suited to all situations, other perhaps than an at least

[10]But see, e.g., [da Costa, 1974], [Grant, 1978], [Priest and Routley, 1984], and [Lin, 1987].

momentary suspension of firm belief in the contradictands when they are initially noticed.[11]

14 Sample problems

Several intuitive problems have motivated us in our work. Among these are the *Brother Problem*, the *Three-wise-men Problem*, the *Examination Problem*, and the *Nell & Dudley Problem*. We briefly describe these and mention their relation to our *reasoning-is-action* thesis.

The Brother Problem: This is due to Moore ([Moore, 1983]) and is as follows: we reason that we do not have a brother, because if we did, we would surely know it. Since we do not know it, we must not have a brother. This is very much like default reasoning (although Moore points out some differences that need not concern us here). The point is, how do we know we do not know we have a brother? If by ''know'' we mean ''know by unbounded reasoning over infinite time'' then we are clearly being unrealistic; yet this is what most formal approaches try to capture. Our hope has been to provide plausible treatments of problems of this sort while retaining up front the temporal character of reasoning-is-action. In Section 18 we illustrate our ideas with a simplified version of this problem tailored to the *as-the-world-turns* approach.

The Three-wise-men Problem: Here three wise men try to draw conclusions based on one another's responses and what each believes about the others' beliefs. See [McCarthy, 1978] for a full discussion. Again traditional formalizations (e.g., [Konolige, 1984] [Kraus and Lehmann, 1987]) assume (in true *stop-the-world* fashion) each wise man to be omniscient, or at least inferentially complete with respect to his internal rules. That is, each wise man already knows all conclusions he can reach in infinite time based on what axioms he starts with. It is more challenging to model the problem according to the *as-the-world-turns* approach. Then an individual wise man is able to reason that another wise man would have been able to deduce by *now* a given conclusion; since he has not *reached* this conclusion, the first wise man is able at the following time step to draw a further conclusion. It is this recognition that reasoning itself takes time which allows the wise man to draw the necessary conclusion. We illustrate this in Section 18.

The Nell & Dudley Problem: Dudley wishes to save Nell, who is tied to the railroad tracks as a train bears down. He must come up with a plan and carry it out before it is too late. Clearly

[11]This latter feature is incorporated into our formalism. See [Elgot-Drapkin and Perlis, 1990]. One obvious approach to explore would be a step-like version of truth-maintenance; see [Doyle, 1979] and [deKleer, 1986].

he must not waste time searching through all possible plans for a theoretical best one. That is, he must take into account the fact that every second he spends planning is one more second gone by and hence one second less in which to carry out a plan---and also one second less before the train reaches Nell.

This latter problem is perhaps the paradigmatic case of those we have discussed, for we have here the deadline scenario in vivid life-and-death form. It also is a case where our approach should be least controversial.

The Examination Problem: This is a similar albeit less dramatic example. A student is taking a timed exam. Initially he spends time planning (i.e., deciding) which part of the exam to work on first. But although this may be very useful and helpful toward improving his overall performance on the exam, it cannot remain so as time goes by. Here again we have a trade-off, and every second spent planning is one second less to actually work on the exam.

Our approach using ‘‘step-logics’’ have been successfully applied to the *Brother Problem* and the *Three Wise-Men Problem* (see [Elgot-Drapkin, 1988], which we return to below.

15 The technical twist: steps

What kind of formal mechanism might lend itself to the *as-the-world-turns* approach? From what we have said, it must include an updated representation of the passage of time, i.e., a clock, which changes its setting as each inference is drawn. This is, after all, what we have been arguing: inference goes on in time, not in a separate land where time is stopped.

For reasons of simplicity and elegance, we have elected to assume there is a fundamental unit or inference time, the so-called *step*. Roughly speaking, an inference takes one step to perform. Of course, complicated reasoning made of many successive inferences in sequence take as many steps as the sequence contains. Thus the reasoner, if it concludes, say, B from A and $A \rightarrow B$, will also conclude that the time when B is concluded (say $t = 11$) is one step (one clock unit) later than the time at which the inference was begun (i.e., $t = 10$).

In intruding time-passage into the formalism itself, we have broken the traditional syntax-semantics barrier. For now the time-symbols have a built-in semantics (given by the system clock: there is a lock-step covariance between them, as long as the software and hardware hold up). This is the primary distinctive feature of step-logics; its use will be amplified on somewhat below (see

our various papers in the bibliography for details).

16 Infinite regress again?

In having the reasoner keep track of time-passage as it reasons, have we not once again made an unwitting argument for keeping track of keeping track, etc? That is, aren't the conclusions as to what time it is (or how many steps have elapsed) themselves taking time, and doesn't our approach commit us to having that passage of time be accounted for too, and so on in an infinite regress? We get around this simply by concluding that it is now step 11 instead of step 10 in parallel with whatever other inferences are going on. Thus we don't require still more time over and above that taken by the ''base'' inference B; both B and $t = 11$ are concluded simultaneously. Note that the recording of time may itself be regarded as a base action, done by an internal physical clock---but this clock action is performed simultaneously with an inference step.

17 Unresolved: dangling time parameters

One technical difficulty that we encountered as we began to study the step approach was as follows: if we want at a given step, say step i, to have the conclusion that the time is now i, we might write this as $t = i$. Now at the next step, this should change to $t = i + 1$. It is easy enough to formalize this (we will see a very simple example in Section 18).[12] But what will happen if we suppose the reasoner to have the rule of *modus ponens* as well as the further belief (conclusion, axiom)[13] that, say, $t = i \rightarrow t = i$? At step i, the inference rule allows an inference to begin, and when it has finished, at step $i + 1$, there will be a resulting conclusion, namely $t = i$! Thus $t = i$ will reappear at step $i + 1$ even though the new time, $i + 1$ is now also registered.

This is just one very simple example of what we call a *dangling time parameter*. The i in the newly concluded wff $t = i$ now ''dangles'' in that it no longer has the appropriate intuitive relation to the actual time, as it did before it was detached from its parent wff $t = i \rightarrow t = i$ at step i. While

[12]Note that this introduces a built-in kind of non-monotonicity, even apart from default situations: truths can go away with the mere passage of time! One way to think of this is that the arrival of the new time information, $t = i + 1$, undoes the old $t = i$. However, this view is not essential to our main point.

[13]We are not particular about terminological details here. What matters is that the reasoner has various pieces of information available to which she can apply her inference rules at any step.

there are kluges around the problem, we have yet to find an elegant one that does not also overly hobble the formalism.[14]

18 Two examples

We have devised a family of formalisms---called *step-logics*---and proven various theorems about their properties. Two step-logics were tailored specifically to the *Brother Problem* and the *Three-wise-men Problem*, respectively. We have implemented these step-logics and successfully applied them to yield real-time solutions (in our sense of *as-the-world-turns*) to these two problems.

As an example of this work, we present a simplified treatment of the *Brother Problem* in step-logic[15].

1 : $\underline{Now(1)}, \underline{(\forall x)[(Now(x) \wedge \neg KB(x-1) \rightarrow \neg B]}$

2 : $\underline{Now(2)}, (\forall x)[(Now(x) \wedge \neg KB(x-1) \rightarrow \neg B], \underline{\neg KB(1)},$

3 : $\underline{Now(3)}, (\forall x)[(Now(x) \wedge \neg KB(x-1) \rightarrow \neg B], \neg KB(1),$
$\underline{\neg B}, \underline{\neg KB(2)}$

Figure 1:

In Figure 1 we see at the left margin so-called *step-numbers* 1, 2, 3, indicating successive times. These times are also recorded internally by the reasoner via the $Now(i)$ wffs, which are seen to change to reflect the successive steps.[16] That is, the reasoner is keeping track of the changing time. The wffs to the right of each step-number are the beliefs of the reasoner at that moment. Underlined wffs are those that have newly arrived at that step, i.e., those that are not simply inherited from the previous step.[17] B is interpreted as the reasoner's having a brother. $KB(i)$ is interpreted as the

[14]One possibility is to insist that tautologies should not be time-dependent, but it is not clear that this is appropriate.

[15]For the full (unsimplified) technical treatment of this problem, as well as more of the general technical approach shoring up the position we advocate, the reader is referred to [Elgot-Drapkin, 1988] and [Elgot-Drapkin and Perlis, 1990].

[16]See [Haas, 1985] for another approach to dealing with the concept of 'Now' --- in his case, by cleverly avoiding it. In our notation, Haas simply rewrites wffs such as $Now(i) \wedge P(i)$ as $P(i)$, and thus eliminates Now altogether. However, this will not do for our purposes. The *as-the-world-turns* reasoner must know that its every reasoning act will alter the ''current'' time, i.e. ''now''.

[17]Inheritance is a complicated matter which we will not go into further here; see [Elgot-Drapkin, 1988] and [Elgot-Drapkin and Perlis, 1990].

reasoner's knowing at step i that she has a brother. She has at step 1 a version of Moore's default axiom to the effect if she does not know at the present time ($x - 1$) that she has a brother then she does not have a brother. She also has two inference rules (not seen): (i) a suitably fancy version of modus ponens, and (ii) negative introspection, enabling her to infer $\neg K B(x - 1)$ at step x if in fact she did not at step $x - 1$ know herself to have a brother. Thus if B is not present at step $x - 1$, $\neg K B(x - 1)$ will appear as a belief at step x.

We see in the figure that $\neg B$ is inferred at step 3, i.e., from having the default belief and the belief (at step 2) that she does not know herself to have a brother, the fancy modus ponens produces the result.[18]

As a second example, we sketch some of the spirit of our solution of the *Three-Wise-Men Problem*, which we present here in the simplified version of two wise men. The sketch unavoidably obscures a number of technical points. Full details for the version with three wise-men are given in [Elgot-Drapkin, 1988] and [Elgot-Drapkin, 1991], where both a formalization and implementation of this problem and its solution are provided.

Suppose then that there are two wise men: W1 and W2. Wise man W1 is our agent, whose reasoning is to be represented by the step-logic. W2 sits behind W1, and a card is placed behind the chair of each, so that W2 can see W1's card but not his own, and W1 can see neither card. Each is to try to figure out the color (black or white) of his own card; the first to do so and state the color correctly is declared the wisest. They are told that at least one of the cards is white. W1 reasons along the following lines, where at each step are shown simultaneous beliefs, and inferences based on them appear at the subsequent step:

1. The time is now 1; if my color were black then W2 would know now that my color is black; W2 knows that if my color is black then his color must be white; if at any moment W2 knows his color then he will say so at the next time step; if at any time W2 makes an utterance, I will know at the next time step that he has done so.

2. The time is now 2; if W2 knew my color were black a step ago then he would have inferred by now that his color is white; and if my color were black then W2 would have known this

[18]This primitive example is not intended to be cognitively plausible, but rather to illustrate the formal apparatus of keeping track of the changing world as one reasons. In [Elgot-Drapkin, 1988] a more sophisticated version of this problem also includes cases of correcting a default conclusion, e.g., retracting the belief that one does not have a brother when evidence to the contrary comes in.

at step 1.

3. The time is now 3; I do not know of an utterance made by W2.
4. The time is now 4; I do not know of an utterance made by W2.
5. The time is now 5; W2 could not have made an utterance at time 3 (since I did not know of such at time 4).
6. The time is now 6; W2 could not have known his color at time 2 (since otherwise he would have made an utterance at time 3).
7. The time is now 7; W2 could not have known my color to be black at time 1 (since he would then have known his color to be white at step 2).
8. The time is now 8; my color cannot be black (since otherwise W2 would have known that at time 1).
9. The time is now 9; my color is white since it is not black.

Notice in our treatment the curious feature of nothing seeming to happen between steps 3 and 4. In fact, something is happening: time is passing and being noted as passing. This in fact is crucial to the solution, for without noting this passage W1 will not be able to show that W2 has had enough time to perform the hypothesized inferences. That is, without, say, the ‘‘delay’’ at time 4, W1 may still conclude at step 4 that W2 did not make an utterance at 2 (instead of 3 as shown in step 5 above) and then at the next step that W2 could not have known his color at time 1 (instead of 2 as shown in step 6 above). Indeed, if we suppose there to be a time 0 when the cards are being set up, this alternate scenario could in part be as follows:

1. (as above)
2. (as above)
3. (as above)
4. The time is now 4; W2 could not have made an utterance at time 2 (since I did not know of such at time 3).

5. The time is now 5; W2 could not have known his color at time 1 (since otherwise he would have made an utterance at time 2).

6. The time is now 6; W2 could not have known my color to be black at time 0 (since he would then have known his color to be white at step 2, not to mention that no cards had at time 0 been set up).

But W1 could not then go on to a further inference at step 7 that W1's color cannot be black, since it is not true that otherwise W2 would have known that at time 0 (due to the cards not having been set up at time 0). W2 could know only by time 1 if W1's card were black, not time 0. Thus the original full nine-step sequence is essential to give W1 enough time (in W1's meta-reasoning about W2's reasoning) to come to a hypothetical conclusion and make a hypothetical utterance, based on a hypothetical assumption that W1's card is black.[19]

This differs markedly from the treatments of the same problem given by [Konolige, 1984] and [Kraus and Lehmann, 1987] in which it is assumed that all results of inferences are simply present in the knowledge base of each agent, rather than introduced over time as the result of an ongoing process of inference. Thus the key issue of W1's judging that enough time has elapsed for W2 to have figured out his own color if W1's color were black, is simply removed from the problem in those treatments.

19 Conclusions and future work

Time, if it is to be treated in a way appropriate to many commonsense settings, must be seen as infecting the entire range of behaviors of agents, including what is perhaps the most salient of all: reasoning. An agent that has no sense that time is passing as it reasons, will not be effective in many common situations.

We have suggested that a non-traditional conceptual framework is needed in order to incorporate reasoning in time about one's very own reasoning in time. A consequence is that a kind of non-monotonicity arises immediately even without defaults, for conclusions do not necessarily inherit from moment to moment, even if one is merely sitting and thinking. If nothing else changes, at

[19]Numerous variations on the details are possible, with the same overall form of behavior; thus it is not essential to suppose a time step passes between utterance and hearing, for example.

least one's belief as to what time it is does: it is 3pm and I am sitting here thinking; now it is 3:01pm. I am not a static entity. I am (part of) where the action is; my reasoning is part of the world's turning.

One approach (step-logics) to such a framework we have explored at length elsewhere---both formally and implementationally---including application to several commonsense problems.

However, our main thesis here is not that any particular formalism has any particular claim to primacy, but rather simply that research in temporal reasoning for intelligent systems has an additional dimension to account for: namely, that reasoning itself takes time, in ways that cannot be ignored by those very systems. We hope to have made a convincing case for this here.

20 Acknowledgements

We would like to thank Sarit Kraus and three anonymous referees for helpful comments.

References

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

[da Costa, 1974] N. da Costa. On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, 15:497--509, 1974.

[Dean and Boddy, 1988] T. Dean and M. Boddy. An analysis of time-dependent planning. In *Proceedings of the 1988 AAAI*, pages 49---54, 1988.

[deKleer, 1986] J. deKleer. An assumption-based TMS. *Artificial Intelligence*, 28:127--162, 1986.

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

[Drapkin and Perlis, 1986a] J. Drapkin and D. Perlis. A preliminary excursion into step-logics. In *Proceedings SIGART International Symposium on Methodologies for Intelligent Systems*, pages 262--269. ACM, 1986. Knoxville, Tennessee.

[Drapkin and Perlis, 1986b] J. Drapkin and D. Perlis. Step-logics: An alternative approach to limited reasoning. In *Proceedings of the European Conf. on Artificial Intelligence*, pages 160--163, 1986. Brighton, England.

[Elgot-Drapkin and Perlis, 1990] 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.

[Elgot-Drapkin *et al.*, 1987] J. Elgot-Drapkin, M. Miller, and D. Perlis. Life on a desert island: Ongoing work on real-time reasoning. In F. M. Brown, editor, *Proceedings of the 1987 Workshop on The Frame Problem*, pages 349--357. Morgan Kaufmann, 1987. Lawrence, Kansas.

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

[Elgot-Drapkin, 1991] J. Elgot-Drapkin. A real-time solution to the wise-men problem. In *Proceedings of the AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, pages 33--40, 1991. Stanford, CA.

[Grant, 1978] J. Grant. Classifications for inconsistent theories. *Notre Dame Journal of Formal Logic*, 19(3), 1978.

[Haas, 1985] A. Haas. Possible events, actual events, and robots. *Computational Intelligence*, 1(2):59--70, 1985.

[Hanks and McDermott, 1986] S. Hanks and D. McDermott. Default reasoning, nonmonotonic logics, and the frame problem. In *Proceedings of the Fifth National Conference on Artificial Intelligence*. AAAI, 1986. Philadelphia, PA.

[Konolige, 1984] K. Konolige. Belief and incompleteness. Technical Report 319, SRI International, 1984.

[Kraus and Lehmann, 1987] S. Kraus and D. Lehmann. Knowledge, belief and time. Technical Report 87-4, Department of Computer Science, Hebrew University, Jerusalem 91904, Israel, April 1987.

[Laird *et al.*, 1987] J. Laird, A. Newell, and P. Rosenbloom. Soar: An architecture for general intelligence. *Artificial Intelligence*, 33(1):1--64, 1987.

[Lin, 1987] F. Lin. Reasoning in the presence of inconsistency. In *Proceedings of the Sixth National Conference on Artificial Intelligence*. AAAI, 1987. Seattle, WA.

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

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

[Moore, 1983] R. Moore. Semantical considerations on nonmonotonic logic. In *Proceedings of the 8th Int'l Joint Conf. on Artificial Intelligence*, 1983. Karlsruhe, West Germany.

[Perlis, 1986] D. Perlis. On the consistency of commonsense reasoning. *Computational Intelligence*, 2:180--190, 1986.

[Perlis, 1988] D. Perlis. Languages with self reference II: Knowledge, belief, and modality. *Artificial Intelligence*, 34:179--212, 1988.

[Priest and Routley, 1984] G. Priest and R. Routley. Introduction: Paraconsistent logics. *Studia Logica*, 43:3--16, 1984.

[Russell and Wefald, 1989] S. Russell and E. Wefald. Principles of metareasoning. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, pages 400---411, 1989.

[Simon, 1978] J. Simon. The limit of computing speed. *Communications of the ACM*, 21(10), 1978.