Reprinted from Journal of Experimental and Theoretical Artificial Intelligence 2(1990)75--98

Reasoning Situated in Time I: Basic Concepts*

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

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

Abstract

The needs of a real-time reasoner situated in an environment may make it appropriate to view error-correction and non-monotonicity as much the same thing. This has led us to formulate situated (or step) logic, an approach to reasoning in which the formalism has a kind of real-time self-reference that affects the course of deduction itself. Here we seek to motivate this as a useful vehicle for exploring certain issues in commonsense reasoning. In particular, a chief drawback of more traditional logics is avoided: from a contradiction we do not have all wffs swamping the (growing) conclusion set. Rather, we seek potentially inconsistent, but nevertheless useful, logics where the real-time self-referential feature allows a direct contradiction to be spotted and corrective action taken, as part of the same system of reasoning. Some specific inference mechanisms for real-time default reasoning are suggested, notably a form of introspection relevant to default reasoning. Special treatment of ''now'' and of contradictions are the main technical devices here. We illustrate this with a computer-implemented real-time solution to R. Moore's *Brother Problem.*

1 Introduction

A resource limitation that is highly evident in commonsense reasoning (i.e., in reasoning about and within a real environment) is simply the passage of time while the reasoner reasons. The paradigm for such a reasoning agent would seem to be that suggested by Nilsson [3], namely, a computer individual with a lifetime of its own. What is of interest for the agent is not 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.

This facilitates the study of *fallible* agents reasoning over time. A fallible agent may derive or encounter an inconsistency, identify it as such, and then proceed to remedy it. Contradictions then need not be bad; indeed, they can be good, in that they allow sources of error to be isolated (see [4]). Of course, contradictions can also be problematic. One desideratum may be that the agent be able to recover from an inconsistent state into a consistent one; we address this ''self-stabilizing'' property in Section 4.

The ''passage of time'' phenomenon is a limitation in the following sense: the conclusions (beliefs) that may be logically (or otherwise) entailed by the agent's earlier beliefs take time to be derived, and time spent in such derivations is concurrent with changes in the world. The issue then is not so much one of weak resources as it is of a real-world fact about processes occurring over time. Indeed, implemented reasoning systems obviously proceed to draw conclusions in steps; see [5, 6, 7, 8].

*The present paper extends ideas presented in [1, 2].

† Supported in part by the IBM Corporation and the U.S. Army Research Office (DAAL03-88-K0087).

‡ Supported in part by the U.S. Army Research Office (DAAL03-88-K0087) and the Martin Marietta Corporation.

The agent should be able to reason *about* its own ongoing reasoning efforts, and in particular, reason whether it has or has not yet reached a given conclusion.[1] One of our main focuses here is the problem of an agent's determining that in fact it does *not* (currently) know something. This *negative introspection* will be a key feature of the deduction, and subsequent resolution, of contradictions in our later examples of default reasoning in Section 6.2. It turns out that negative introspection presents certain temporal constraints that will strongly influence the formal development.

Traditional approaches to formalizing commonsense reasoning suffer from the problem of logical omniscience: if an agent has $\alpha_1, \ldots, \alpha_n$ in its belief set, and if γ is logically entailed by $\alpha_1, \ldots, \alpha_n$, then the agent also believes γ. As a specific example, if an omniscient agent believes α, and also believes $\alpha \rightarrow \beta$, then the agent believes β. As an illustration, refer to Figure 1. The reasoner begins with a set of axioms, and the deductive mechanism generates theorems along the way, e.g., α, later $\alpha \rightarrow \beta$, still later β. Such mechanisms have usually been studied in terms of the set of all theorems deducible therein, what we call the ''final tray of conclusions'' into which individually proven theorems are represented as dropping, thereby ignoring their time and means of deduction. One asks, for instance, whether a wff α is a theorem (i.e., is in the final tray), *not* whether α is a theorem proven in i steps.

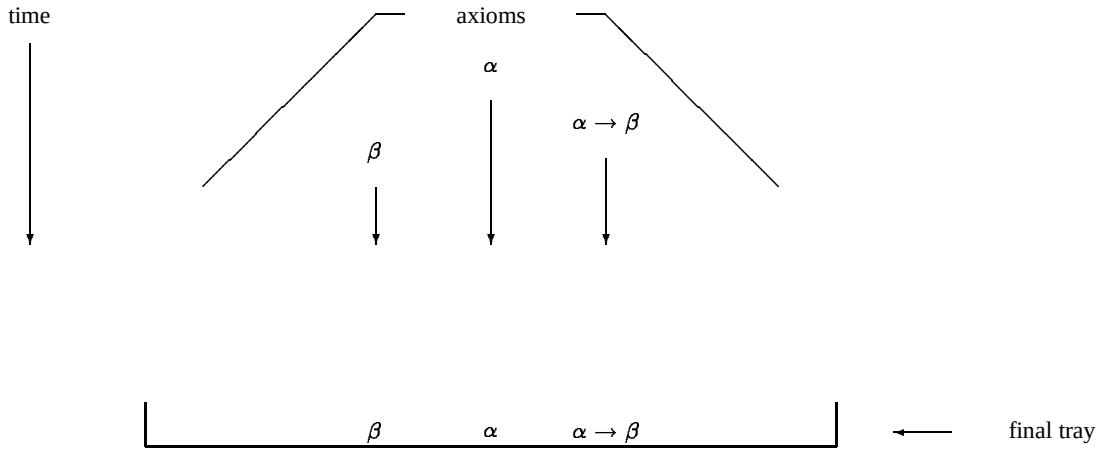

Figure 1: Final-tray logical studies

A particularly vexing aspect of this type of reasoning is what we call the *swamping problem*---namely that from a contradiction all wffs are concluded. For this reason most formal studies of reasoning deliberately avoid contradictions; those that do not (e.g., Doyle [9]), provide a separate device for noting contradictions and revising beliefs while the ''main'' reasoning engine sits quiescent. In general, however, this will not do, since the knowledge needed to resolve conflicts will depend on the same wealth of world knowledge used in any other reasoning. Thus reasoning about birds involves inference rules applied to beliefs about birds, whether used to resolve a conflict or simply to produce non-conflicting conclusions. We contend, then, that one and the same on-going process of reasoning should be responsible both for keeping itself apprised of contradictions and their resolution, and for other forms of reasoning.

The literature contains a number of approaches to limited (non-omniscient) reasoning, apparently with similar motivation as our own. However, with very little exception, the idealization 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 [10] 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 [11] and Fagin and Halpern [12] 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

[1] This separates two directions for study. First, one would like a meta-theory 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) about what it has and has not done at any given time. These are obviously interrelated, and yet can be tackled somewhat independently. This is taken up below.

are not explicit. Fagin and Halpern in particular postulate a notion of awareness, so that if α and $\alpha \rightarrow \beta$ are known, still β will not be concluded unless the agent is aware of β; just how it is that β fails to be in the awareness set in unclear. Our own approach provides a rather different notion of awareness, where the agent is aware of all closed sub-formulas of its beliefs; hence the awareness set changes over time. Goodwin [5] comes a little closer to meeting our desiderata but still maintains a largely final-tray-like perspective.

In what follows we will outline our own suggestion for formalizing commonsense reasoning in a way that seems amenable to real-time issues. Section 2 presents the underlying idea, which we call *step-logic*. Section 3 gives some more details, and Section 4 presents several technical definitions and results for step-logics. In Section 5 we discuss the most primitive step-logic, and in Section 6 we present a much more sophisticated one. Both have been implemented in PROLOG. In Section 7 we discuss difficulties with representing ''now'' and with handling contradictions.

We expect further research in step-logic to lead to insights into a number of issues in commonsense reasoning. In particular, we think that step-logic is a natural setting for real-time versions of reasoning appropriate for problems such as the *Gun problem*,[2] and the *Three-wise-men problem*, in which each wise man reasons about the time available vis-a-vis the others' thought processes.[3]

One further commonsense problem is the *Nell and Dudley problem*, in which Dudley must save Nell from an onrushing train. Clearly he must formulate and carry out a plan of action quickly, and must take into account that every extra second spent planning leaves that much less time for acting. The fact that ''now'' changes as one thinks is particularly germane in this problem. The present paper will not offer a solution to this problem; it is currently under investigation. However, we have designed step-logic with an eye to this problem, and thus we have made the concept of ''now'' an integral part of the formal treatment.

2 A Time-situated View

Ordinary logic serves well the purpose of modelling a reasoning agent's activity from *afar*, as a meta-theory about the agent. This is a useful thing; still, it is also of interest to have a direct representation of the evolving process of the very reasoning itself. This can be done in ordinary logic if the representation is in the meta-theory, say by means of a time argument to a predicate representing the agent's proof process. Indeed, the most elementary step-logic we have proposed is just of this sort. However, in order for the agent to reason about the passage of time that occurs *as* it reasons, time arguments must be put into the agent's own language. That is, such an agent's logic (a step-logic) would evolve and represent that evolving history at the same time. Can this be anything at all like a traditional logic? It can, and not merely by implementing a deductive engine and watching it go through states one by one. This is not so very surprising, for this seems to be what humans do: we are constantly going on in time, and yet reasoning in time, even reasoning about time as we go on in time.

There will be salient differences from ordinary logic, however. Since time goes on as the agent reasons, and since this phenomenon is part of what is to be reasoned about, the agent will need to take note of facts that come and go, e.g., ''It is now 3pm and I am just starting this task . . . Now it is no longer 3pm, but rather it is 3:15pm, and I still have not finished the task I began at 3pm.'' So, as time (and the agent's reasoning) goes on, the former conclusion that ''It is now 3pm'' needs to be retracted, in favor of the new conclusion ''It is now 3:15pm''. This immediately puts us in a non-traditional setting, for we lose monotonicity: as the history evolves, conclusions may be lost.[4] Their loss, however, need not be considered a weakness, but rather a strength, based on a reasoned assessment of a changing situation. It is clear, then, that a step-logic cannot in general retain or inherit all conclusions from one step to the next. We caution the reader to keep this in mind in our examples. Despite this feature, we will see that step-logic is primarily a deductive apparatus.[5]

[2] See Hanks and McDermott [13]. In part we see the approach we will advocate below as lying mid-way between their position and the logicist tradition they criticize.

[3] See [14] and [15] for various descriptions of this problem and its final-tray-like solutions, and [16] for a solution using step-logic.

[4] That is, the new information that ''it is now 3:15pm'' can be thought of as erasing the old information that ''it is now 3pm''. While this is not strictly non-monotonic in the usual sense, it has a similar flavor.

[5] To be sure, non-monotonic formalisms already exist in the literature [17, 18, 19]. However, they do not explicitly treat on-going processes in the reasoning modelled. We suspect that the finely honed non-monotonicities in those studies may amount to a kind of temporal reasoning that would be brought out if they were applied to problems in which an agent comes across conflicts; see [4].

The issue of representing time within a logic has been studied intensively, e.g., by Allen [20], McDermott [21], and McKenzie and Snodgrass [22]. However, such representations of time are not related in any obvious way to the process of actually producing theorems in that *same* logic. In effect, we want to augment logic with a notion of ''now'', which appropriately changes as deductions are performed. It turns out that this is not an easy task. While there are many issues related to the general approach we are advocating, we will concentrate here on describing some useful technical devices in time-situated reasoning that pertain to negative introspection.

In contrast to already existent approaches, we propose step-logic to model reasoning that focuses on the on-going process of deduction; see Figure 2. The reasoner starts out with an empty set of beliefs at time 0. Certain ''conclusions'' or ''observations'' may arise at discrete time steps. At some time, i, it may have belief α, concluded based on earlier beliefs, or as an observation arising at step i. At some later time, j, it comes up with $\alpha \rightarrow \beta$. Later still, the agent might deduce β. Of course, much the same might be said of any deductive logic. However in step-logic these time parameters can figure in the on-going reasoning itself. The rest of the paper is devoted to describing some details and uses of this phenomenon.

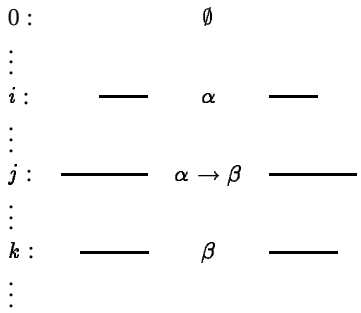

Figure 2: Step-like logical studies

3 The Basics

A step-logic is characterized by a language, observations, and inference rules. We emphasize that step-logic is *deterministic* in that at each step i *all* possible conclusions from the rules of inference applied to the previous step are drawn (and therefore are among the wffs at step i). However, for real-time effectiveness and cognitive plausibility, at each step we want only a finite number of conclusions to be drawn.[6] Thus a number of issues present themselves: smallness of the conclusion set at each step, contradiction-handling, etc. In this paper we are principally concerned with studying default reasoning in terms of contradiction-handling. That is, *we want step-logic to allow contradictions to arise* as defaults are invoked and yet also *we want the consequences of contradictions to be controlled* in a reasonable manner consistent with commonsense.

We return now to the idea that there are two distinct types of formalisms of interest, that occur in pairs: the meta-theory SL^n *about* an agent, and the agent-theory SL_n itself. Here n is simply an index serving to distinguish different versions of step-logics. It is the latter, SL_n, that is to be step-like; the former, SL^n, is simply our assurance that we have been honest in describing what we mean by a particular agent's reasoning. Thus the meta-theory is to be a scientific theory subject to the usual strictures such as consistency and completeness. The agent theory, on the other hand, may be inconsistent and incomplete; indeed if the agent is an ordinary fallible reasoner it *will* be so. The two theories together form a step-logic *pair*.

[6] Indeed it should be not just finite, but small. Our current idealization does not go this far; we intend, however, to eventually make broad use of a ''retraction'' mechanism to keep things to a reasonable size. Specifically, we anticipate the introduction of a notion of relevance, along the lines we pursued in [6, 7, 8] for a computational model of memory; in fact, our entire approach can be viewed as a formalization and abstraction of our memory-model work. However, the present paper makes only a modest use of retraction (see Section 6.1).

We propose three major mechanisms to study as possible aspects of an agent-theory: self-knowledge, time, and retraction. Since it is important for the agent to reason about its own processes, a self, or belief, predicate is needed. We employ a predicate symbol, K, for this purpose: $K(i, \text{`}\alpha\text{'})$ is intended to mean that the agent knows wff α at time i.[7] K may or may not be part of the agent's own language; however, many kinds of reasoning require that it be. Note that 'α' is a name for α, i.e., a constant term.[8] We drop the quotes in $K(i, \text{`}\alpha\text{'})$ in the remainder of the paper.

In order for the agent to reason about time, a time predicate is needed. This not only amounts to a parameter such as i in $K(i, \alpha)$ as we just saw, but information as to how i relates to the on-going time as deductions are performed. Thus the agent should have information as to what time it is *now*, and this should change as deductions are performed. We use the predicate expression $Now(i)$ to mean the time currently is i. Again, this may or may not be part of the agent's language, but in many cases of interest it is.

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 may be 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 need some kind of a retraction device. Retraction will be facilitated by focusing on the dual: inheritance. We do *not* assume that all deductions at time i are inherited (retained) at time $i + 1$. By carefully restricting inheritance we achieve a rudimentary kind of retraction. The most obvious case is that of $Now(i)$. If at a given step the agent knows the time to be i, by having the belief $Now(i)$, then that belief shall not be inherited to the next time step.

Here we encounter a general phenomenon of temporal constraint that will pervade the rest of our development. Consider the process of concluding by default, on the basis of not knowing X ''now,'' that X is false (where X is any assertion, possibly dependent on time). But how, at time i, can an agent determine that it does not know X at time i? Intuitively, certain beliefs have accumulated at time i, and only *then* can the further belief be formed, that X is not among the former. Thus the negative introspective conclusion seems to come *after* the time at which X is in fact not present: it is concluded, say, at time $i + 1$, that X was not known at i. Now this introspective time-delay may seem to be a mere quibble; but if we ignore it, trouble arises. For suppose that we write the above default as follows:

$$(\forall t)[(Now(t) \wedge \neg K(t, X)) \rightarrow \neg X]$$

That is, if we don't currently know X, then conclude $\neg X$. If this is one of the beliefs present at time i, and if the beliefs $Now(i)$ and $\neg K(i, X)$ are also present at time i, then indeed the conclusion $\neg X$ may be derived by some appropriate rule of inference in the next step, $i + 1$. But now, let's consider the belief $\neg K(i, X)$. This appears (through negative introspection) in the set of beliefs at time i, on the basis of X *not* being in that same set. There are problems with this, for we then are not really dealing with a fixed set for time i, but rather a two-stage production in which beliefs are gathered initially and then an introspective process is allowed to add to that set, playing fast and loose with the meaning of ''not being known at time i.'' This in turn leads to severe ambiguities, in that the very process of inserting, say, $\neg K(i, X)$ into the beliefs at time i results in something being known after all, something that was *not* really known at time i, namely $\neg K(i, X)$ itself.

But suppose we grant that some oracle manages to place all negative introspective conclusions *about* the time-i belief set *into that very same set*. This unfortunately forces an infinite set of beliefs into that set, since there are infinitely many unknown formulas at any step. Yet our approach of real-time reasoning commits us to a finite belief set at all steps. Thus we must forego the luxury of having the agent be able to know that it doesn't know a given fact *now*; instead the best that can be done is to know that it didn't know the fact a moment ago, when it last was able to scan its belief set. The act of scanning has changed the world, at least in the sense that it has taken time. Thus the agent's self-knowledge lags slightly behind. We then will represent the above default reasoning in the following altered form:

$$(\forall t)[(Now(t) \wedge \neg K(t - 1, X)) \rightarrow \neg X]$$

If we didn't know X a moment ago, then conclude $\neg X$. Suppose this is a belief at time i. If at time $i - 1$, we did not have the belief X, then, using the revised notion of introspection, at time i we can negatively introspect to produce

[7] We are not distinguishing here between belief and knowledge. See [23] for a discussion of belief vs. knowledge.

[8] In this paper we do not address the case of α having free variables in $K(i, \text{`}\alpha\text{'})$.

the belief $\neg K(i-1, X)$. If we also have the belief $Now(i)$ at time i, a suitable form of *modus ponens* allows us to conclude $\neg X$ at time $i+1$.

Aside from obvious real-time relevance, the Now predicate is important in other ways. For instance, above we illustrated its use in representing default information; more will appear on this later, in Section 6.2. The *Three-wise-men problem* involves drawing the conclusion that one has a white spot on the basis of the behavior of others over time, and in particular on how much time has elapsed; proposed solutions that do not use something like a Now predicate assume omniscience of the reasoners and thus lose much of the sense of the original problem. Finally, the *Nell and Dudley problem* requires a changing time so that Dudley will be able to recognize when it has become too late to do anything.

In [1] we proposed eight step-logic pairs, arranged in increasing sophistication, with respect to the three mechanisms above (self-knowledge, time, and retraction). In our current notation, these are $< SL_0, SL^0 >, \ldots, < SL_7, SL^7 >$. SL_0 has none of the three mechanisms, and SL_7 has all. Of the eight agent-theory/meta-theory pairs, only SL^0 and SL_7, the simplest meta-theory and the most complex agent-theory, have been studied in any detail.[9] The meta-theories all are consistent, first-order theories, and therefore complete with respect to standard first-order semantics. However, their associated agent-theories are another matter. These we do not even *want* in general to be consistent, for they are (largely) intended as formal counterparts of the reasoning of fallible agents. SL_0 is an exception, for it, as an initial effort, was constructed to do merely propositional (tautological) reasoning so we could more easily test its meta-theory, SL^0.

A notion of completeness for the meta-theory is defined as follows:

Definition .1 *A meta-theory SL^n is* analytically complete, *if for every positive integer i, and every constant α naming an agent wff of the corresponding agent-theory, either $SL^n \vdash K(i, \alpha)$ or $SL^n \vdash \neg K(i, \alpha)$.*[10]

We showed that our SL^0 formalism is in fact analytically complete. But what kind of completeness might be wanted for an *agent* theory? In SL_0, it is desirable that every tautology be (eventually) provable. This is the case, since every tautology has a proof in propositional logic and, for a sufficiently large value of i, all axioms (i.e., the ''observations'') in such a proof will have appeared (by design of SL_0) by step i. Thus SL_0 is complete with respect to the intended domain, namely, tautologies. However, for other step-logics the case is not so simple, for the intended domain, namely, the commonsense world, has no well-understood precise definition. Nevertheless, we can isolate special cases in which certain meta-theorems are possible. In particular, if no non-logical axioms (beliefs) are given to an agent at step 0 (or any later time), then it is reasonable to expect the agent to remain consistent. This we will be able to establish for all our agent logics in which the logical axioms do not contain the predicate symbol ''Now''.

4 Definitions and Theorems

We now present several definitions, most of which are analogous to standard definitions from first-order logic. Consequently certain results follow trivially from their first-order counterparts.

Intuitively, we view an agent as an inference mechanism that may be given external inputs or observations. Inferred wffs are called beliefs; these may include certain observations.

Let $\mathcal{L}$ be a first-order language, and let $\mathcal{W}$ be the set of wffs of $\mathcal{L}$.

Definition .2 *An* observation-function *is a function $OBS : \mathbf{N} \rightarrow \mathcal{P}(\mathcal{W})$, where $\mathcal{P}(\mathcal{W})$ is the powerset of $\mathcal{W}$, and where for each $i \in \mathbf{N}$, the set $OBS(i)$ is finite. If $\alpha \in OBS(i)$, then α is called an* i-observation.

Definition .3 *A* history *is a finite tuple of pairs of finite subsets of $\mathcal{W}$. $\mathcal{H}$ is the class of all histories.*

Definition .4 *An* inference-function *is a function $INF : \mathcal{H} \rightarrow \mathcal{P}(W)$, where for each $h \in \mathcal{H}$, $INF(h)$ is finite.*

Intuitively, a history is a conceivable temporal sequence of belief-set/observation-set pairs. The history is a *finite* tuple; it represents the temporal sequence up to a certain point in time. H consists of all conceivable histories, not

[9] We describe SL^0 in Section 5 and SL_7 in Section 6.

[10] K then has two roles: in SL^n as used here, and in SL_n. The context will make the role clear.

merely those that occur in some actual course of reasoning. The inference-function extends the temporal sequence of belief sets by one more step beyond the history. Figure 3 illustrates one such observation-function and inference-function. We can see that INF depends both on OBS and the histories, and that any given history depends both on OBS and INF. We have illustrated one such history: the history of the first five steps.[11] Definitions .5 and .6 formalize these concepts in terms of a step-logic SL_n.

Let

- $OBS(i) = \begin{cases} \{bird(x) \rightarrow flies(x)\} & \text{if } i = 1 \\ \{bird(tweety)\} & \text{if } i = 3 \\ \emptyset & \text{otherwise} \end{cases}$
- $Thm_i \subseteq \mathcal{W}$, $0 \leq i < n$; $Thm_0 = \emptyset$;
- $INF(<<Thm_0, OBS(1)>, \ldots, <Thm_{n-1}, OBS(n)>>) =$
  $Thm_{n-1} \cup OBS(n) \cup \{\alpha(t) \mid (\exists\beta)(\beta(t), \beta(x) \rightarrow \alpha(x) \in (Thm_{n-1} \cup OBS(n)))\}$.

The history h of the first five steps then would be:

$$\begin{array}{lcccc} h = & << & \emptyset & , \{bird(x) \rightarrow flies(x)\} & > , \\ & < & \{bird(x) \rightarrow flies(x)\} & , \emptyset & > , \\ & < & \{bird(x) \rightarrow flies(x)\} & , \{bird(tweety)\} & > , \\ & < & \{bird(x) \rightarrow flies(x), bird(tweety), flies(tweety)\} & , \emptyset & > , \\ & < & \{bird(x) \rightarrow flies(x), bird(tweety), flies(tweety)\} & , \emptyset & >> \end{array}$$

Figure 3: Example of a particular OBS and INF

Definition .5 *An* SL_n-theory over a language $\mathcal{L}$ *is a triple,* $< \mathcal{L}, OBS, INF >$*, where* $\mathcal{L}$ *is a first-order language,* OBS *is an observation-function, and* INF *is an inference-function. We use the notation,* $SL_n(OBS, INF)$*, for such a theory (the language* $\mathcal{L}$ *is implicit in the definitions of* OBS *and* INF*). If we wish to consider a fixed* INF *but varied* OBS*, we write* $SL_n(\cdot, INF)$*.*

Let $SL_n(OBS, INF)$ be an SL_n-theory over $\mathcal{L}$.

Definition .6 *Let the set of 0-theorems, denoted* Thm_0*, be empty. For* $i > 0$*, let the set of i-theorems, denoted* Thm_i*, be* $INF(<<Thm_0, OBS(1)>, <Thm_1, OBS(2)>, \ldots, <Thm_{i-1}, OBS(i)>>)$*. We write* $SL_n(OBS, INF) \vdash_i \alpha$ *to mean* α *is an i-theorem of* $SL_n(OBS, INF)$*.*[12]

Definition .7 *Given a theory* $SL_n(OBS, INF)$*, a corresponding* SL^n-theory*, written* $SL^n(OBS, INF)$*, is a first-order theory having binary predicate symbol* K*,*[13] *numerals, and names for the wffs in* $\mathcal{L}$*, such that*

$$SL^n(OBS, INF) \vdash K(i, \alpha) \text{ iff } SL_n(OBS, INF) \vdash_i \alpha.$$

Thus in $SL^n(OBS, INF)$, $K(i, \alpha)$ is intended to express that α is an i-theorem of $SL_n(OBS, INF)$.[14]

Let $\mathcal{L}'$ be the language having the symbols of $\mathcal{L}$ and the (possibly additional) predicate symbols K and Now. Thus $\mathcal{L}'$ may be $\mathcal{L}$ itself.

Definition .8 *A* step-interpretation *for* $\mathcal{L}'$ *is a sequence* $M =< M_0, M_1, \ldots, M_i, \ldots >$*, where*

1. *Each* M_i *is an ordinary first-order interpretation of* $\mathcal{L}'$*.*
2. $M_i \models Now(i)$*.*

[11] This example serves to illustrate how these three concepts are inter-related. There are many possibilities for defining the functions OBS and INF; hence, many different histories are possible.

[12] Note the non-standard use of the turnstile here.

[13] We see that the predicate letter K has two roles: in SL^n and in SL_n. The context will make the role clear.

[14] In [1, 2] we used $K(i, \alpha)$ for $K(i, \alpha)$.

Definition .9 *A* step-model *for* $SL_n(OBS, INF)$ *is a step-interpretation* M *satisfying*

1. $M_i \models K(j, \alpha)$ *iff* $SL_n(OBS, INF) \vdash_j \alpha$.
2. $M_i \models \alpha$ *whenever* $SL_n(OBS, INF) \vdash_i \alpha$.

Condition 1 insures that a chronological record of the j-theorems exists in each M_i; and Condition 2 insures that the i-theorems are in fact true. M should not be thought of as the real external world, corresponding to an agent's beliefs. Rather, M is just a reflection of those beliefs and may or may not correspond to external matters. In particular, a wff B can be true in M_i and false in M_{i+1} simply because the agent has changed its mind.

Definition .10 *A wff* α *is* i-true *in a step-model* M *(written* $M \models_i \alpha$*) if* $M_i \models \alpha$.

Definition .11 $SL_n(OBS, INF)$ *is* step-wise consistent *if for each* $i \in \mathbf{N}$*, the set of* i*-theorems is consistent (classically, i.e., the set has a first-order model).*

Definition .12 $SL_n(OBS, INF)$ *is* eventually consistent *if* $\exists i$ *such that* $\forall j > i$*, the set of* j*-theorems is consistent.*

Definition .13 *An observation-function OBS is* finite *if* $\exists i$ *such that* $\forall j > i,\ OBS(j) = \emptyset$.

Definition .14 $SL_n(\cdot, INF)$ *is* self-stabilizing *if for every finite OBS,* $SL_n(OBS, INF)$ *is eventually consistent.*

Remark .15
1. *Even if* $SL_n(OBS, INF)$ *is step-wise consistent, it can have conflicting wffs at* different *steps, e.g.,*
   $SL_n(OBS, INF) \vdash_{10} Now(10)$ *and* $SL_n(OBS, INF) \vdash_{11} \neg Now(10)$.
2. *Any step-wise consistent theory is eventually consistent.*
3. *Intuitively a self-stabilizing theory* $SL_n(\cdot, INF)$ *corresponds to a fixed agent that can regain and retain consistency after being given arbitrarily (but finitely) many contradictory initial beliefs.*

Theorem .16 *If* $SL_n(OBS, INF)$ *has a step-model, then it is step-wise consistent.*[15]

Proof: Let $SL_n(OBS, INF)$ have a step-model $M = < M_0, M_1, \ldots, M_i, \ldots >$. Let $j \in \mathbf{N}$ be arbitrary. Then for each α in the set of j-theorems, $M_j \models \alpha$. This means that the set of j-theorems is consistent, since it has a (standard first-order) model M_j. □

Theorem .17 (Soundness) *Every step-logic* $SL_n(OBS, INF)$ *is sound with respect to step-models. That is, every* i*-theorem* α *of* $SL_n(OBS, INF)$ *is* i*-true in every step-model* M *of* $SL_n(OBS, INF)$*, i.e.,*
if $SL_n(OBS, INF) \vdash_i \alpha$ *then* $M \models_i \alpha$.

Proof: Let α be an i-theorem of $SL_n(OBS, INF)$, and let M be a step-model of $SL_n(OBS, INF)$.
$SL_n(OBS, INF) \vdash_i \alpha$, so by definition of step-model, $M_i \models \alpha$, and hence (by definition of i-true) $M \models_i \alpha$. □

5 SL_0 and SL^0

The first step-logic pair we investigated was $< SL_0, SL^0 >$. The language of SL_0 is propositional, where the propositional letters are $P_0, P_1, P_2, \ldots$. The meta-theory SL^0 is a first-order theory as described in Definition .7. SL_0 corresponds to the reasoning of a very simple agent that can deduce only tautologies. The agent is ''fed'' beliefs (its ''observations'') consisting of special tautologies, from which it is to draw others. In [24] we formalized the meta-theory SL^0 for describing the steps taken by such an agent.[16]

[15] This result will be useful in showing certain step-logics are consistent; however, by the same token, since many interesting step-logics are *inconsistent* (and in fact derive much of their interest from their inconsistency), step-models are not sufficiently general as defined. We intend to explore a broader concept of step-model in future work.

[16] Although there we did not yet use the notational distinction of SL_0 and SL^0.

To have the agent deduce all tautologies, it is necessary to provide sufficiently many axioms. The usual approaches involve schemata encoding an infinite number of axioms (see [25]), yet we wish the agent to have only a finite number of beliefs at each step. To achieve this, we ''feed in'' first-order logical axioms little by little (according to increasing bounds on their lengths (i.e. the number of connectives) and ranges of symbols used) through the observation-function. That is, an instance α of an axiom schema is an i-observation iff the length of α and the highest index j of any propositional letter P_j in α are both less than i. For example, $P_0 \rightarrow (P_0 \rightarrow P_0)$ is a 3-theorem, but is not a 0-,1-, or 2-theorem. Although the highest index of this wff is zero, it has a length of two, and is therefore not ''fed in'' until step 3.

Theorem .18 *SL^0 is analytically complete.*

The proof is a long series of lemmas involving induction on the length of formulas. See [24] for the complete proof.

SL^0 was studied to gain an understanding of the underlying idea of step-logic, and to gain some practical experience.[17] Although SL^0 was studied in some detail, SL_0 is not an appropriate step-logic for commonsense reasoning: not only is the propositional language too weak, but an arbitrarily large number of tautologies are fed in at each step. A commonsense reasoner should have only a relatively small number of active beliefs with which to work at each step.[18]

6 SL_7

In this section we outline what is so far the most ambitious step-logic: SL_7.[19] SL_7, as stated earlier, is *not* intended in general to be consistent. If supplied *only* with logically valid wffs that are Now-free on which to base its reasoning, then indeed SL_7 will remain consistent over time: there will be no step i at which the conclusion set is inconsistent, for its rules of inference are sound (see Theorem .22 in Section 6.1). However, virtually all the interesting applications of SL_7 involve providing the agent with some non-logical and potentially false axioms, thus opening the way to derivation of contradictions. This behavior is what we are interested in studying, in a way that avoids the swamping problem. The controlled growth of deductions in step-logic provides a convenient tool for this, as we will see.

Section 6.1 provides some details of SL_7, and in Section 6.2 we use Moore's *Brother problem* as an illustration.

6.1 The Details

The language of SL_7 is first-order, having unary predicate symbol, Now, binary predicate symbol, K, and ternary predicate symbol, $Contra$, for time, knowledge, and contradiction, respectively. We write $Now(i)$ to mean the time is now i; $K(i, \alpha)$ can be thought of as stating that α is known [20] at step i; and $Contra(i, \alpha, \beta)$ means that α and β are in direct contradiction (one is the negation of the other) and both are i-theorems.

The formulas that the agent has at step i (the i-theorems) are precisely all those that can be deduced from step $i-1$ using the applicable rules of inference. As previously stated, the agent is to have only a finite number of theorems (conclusions, beliefs, or simply wffs) at any given step. We write:

$$\begin{aligned} i: &\quad \ldots, \alpha \\ i+1: &\quad \ldots, \beta \end{aligned}$$

[17] An implementation of SL^0 has been written in PROLOG, and was run on an IBM PC-AT.

[18] This failing of SL_0 can be seen in our implementation, where at a very early step so many theorems have accumulated that their computation on an IBM PC-AT is severely hindered.

[19] The earlier SL_i's are weaker versions, missing either time or retraction or belief/knowledge predicates, and therefore too weak for our purposes in this paper. Also, SL_7, unlike SL_0, is intended not for derivation of tautologies but rather for the study of particular default capabilities; in particular, tautologies and other logical axioms are not generally employed in SL_7. Finally, we use the notation SL_7 for any of a family of step-logics whose OBS and INF involve the predicates Now and K and contain a retraction mechanism. Choosing OBS and INF therefore fixes the theory within the family.

[20] known, believed, or concluded. The distinctions between these (see [23, 4, 26]) will not be addressed here.

to mean that α is an i-theorem, and β is an $i+1$-theorem. There is no implicit assumption that α (or any other wff other than β) is present (or not present) at step $i+1$. The ellipsis simply indicates that there might be other wffs present. Wffs are not assumed to be inherited or retained in passing from one step to the next, unless explicitly stated in an inference rule. In Figure 4 below, we illustrate one possible inference function, denoted INF_B, involving a rule for special types of inheritance; see Rule 7.

For *time*, we envision a clock which is ticking as the agent is reasoning. At each step in its reasoning, the agent looks at this clock to obtain the time.[21] The wff $Now(i)$ is an i-theorem. $Now(i)$ corresponds intuitively to the statement ''The time is now i.''

Self-knowledge involves the predicate K, and (in INF_B) a new rule of inference, namely a rule of (negative) introspection; see Rule 5 in Figure 4 below. This rule is intended to have the following effect. $\neg K(i, \alpha)$ is to be deduced at step $i+1$ if α is not an i-theorem, but does appear as a closed sub-formula at step i.[22] We regard the closed sub-formulas at step i as approximating the wffs that the agent is ''aware of'' at i.[23] Thus the idea is that the agent can tell at $i+1$ that a given wff it is *aware* of at step i is not one of those it has as a *conclusion* at i. See [12] for another treatment of awareness. We will use the K concept to allow the agent to negatively introspect, i.e., to reason at step $i+1$ that it did not know β at step i. Thus, using INF_B, if α and $\alpha \rightarrow \beta$ are i-theorems, then β and $\neg K(i, \beta)$ will be $i+1$-theorems (concluded via Rules 3 and 5, respectively). Currently we do not employ positive introspection (i.e., from α at i infer $K(i, \alpha)$ at $i+1$), although it can be recaptured from axioms if needed.

Retractions are used to facilitate removal of certain conflicting data. Handling contradictions in a system of this sort can be quite tricky. Currently we handle contradictions by simply not inheriting the formulas directly involved.[24] Unlike SL_0 which is monotonic (that is, if α is an i-theorem, then α is also an $i+1$-theorem), SL_7 is non-monotonic. In $SL_7(\cdot, INF_B)$, a conclusion in a given step, i, is inherited to step $i+1$ if it is not contradicted at step i and it is not the predicate $Now(j)$, for some j; see Rule 7 in Figure 4 below.

We formulated $SL_7(\cdot, INF_B)$ with applications such as the *Brother problem* (see Section 6.2) in mind. This led us to the rules of inference listed in Figure 4. Rule 3 states, for instance, that if α and $\alpha \rightarrow \beta$ are i-theorems, then β will be an $i+1$-theorem. Rule 3 makes no claim about whether or not α and/or $\alpha \rightarrow \beta$ are $i+1$-theorems. The axioms (i.e., the ''observations'') are those listed in Section 6.2.

Note that central to our approach is the idea that, for at least some conclusions that our agent is to make, the time the conclusion is drawn is important. For instance, if it concluded at time (step) 5 that some wff B is unknown, we prefer the agent to conclude $\neg K(5, B)$ rather than simply $\neg K(B)$. The reason for this is that it may indeed be true that B is unknown at time 5, but that later B becomes known; this latter event however should not force the agent to forget the (still true) fact that *at time* 5, B was unknown. On the other hand, if we put time stamps on *all* conclusions, then B itself, once concluded, will require more complex inheritances in order to carry B on from step to step as a continuing truth. Thus it seems preferable not to time-stamp every conclusion. This leaves us with a problem of deciding which conclusions to stamp; currently we are stamping only introspections, contradictions, and ''clock look-ups''.

It is worth amplifying on the use of Contra. Suppose that at step i the agent has the wffs $\neg\alpha$, $\neg\beta$, and $\alpha \vee \beta$. (They are all i-theorems.) While these are indeed mutually inconsistent, they do not form a *direct* contradiction; it takes some further work to see the contradiction. If, for instance, at step $i+1$ the agent deduces β (say, from a further wff $\neg\alpha \wedge (\alpha \vee \beta) \rightarrow \beta$ also present at step i), then at step $i+1$ there would be a direct contradiction. This would then be noticed (via Rule 6) at step $i+2$ with the wff $Contra(i+1, \beta, \neg\beta)$. Then (by Rule 7) neither β nor $\neg\beta$ would be inherited to step $i+3$. Note that what is not inherited is context-dependent: if a slightly different line of reasoning had led from the same wffs at step i to a different contradiction at $i+1$, different wffs would fail to be inherited. Thus it is the actual time-trace of past reasoning that is reflected in the decision as to what wffs to distrust. Also note that if the extra wff that allowed the implicit contradiction to become direct had not been present, the implicit contradiction

[21] Richard Weyhrauch analyzed this idea in a rather different way in his talk at the Sardinia Workshop on Meta-Architectures and Reflection, 1986; see [27].

[22] A sub-formula of a wff is any consecutive portion of the wff that itself is a wff. Note that there are only finitely many such sub-formulas at any given step. Rule 5 formalizes the introspective time-delay discussed in Section 3.

[23] ''You can't know you don't know something, if you never heard of it.'' Thus from beliefs $Bird(x) \rightarrow Flies(x)$ and $Bird(tweety)$ at step i, $Bird(tweety) \rightarrow Flies(tweety)$ may follow at step $i+1$. Then at step $i+1$, Flies(tweety) would become something the agent is aware of. (In INF_B this will certainly be the case, and in fact $Flies(tweety)$ will even be a theorem.)

[24] In future work we hope to have a mechanism for tracing the antecedents and consequents of a formula α when α is suspect, a la Doyle and deKleer (see [28, 29]), though in the context of a real-time reasoner.

The inference rules given here correspond to an inference-function, INF_B. For any given history, INF_B returns the set of all immediate consequences of Rules 1--7 applied to the last step in that history. Note that Rule 5 is the only default rule.

Rule 1 :	$\dfrac{i : \ldots}{i+1 : \ldots, Now(i+1)}$	Corresponds to looking at clock
Rule 2 :	$\dfrac{i : \ldots}{i+1 : \ldots, \alpha}$	If $\alpha \in OBS(i+1)$---Obs. become beliefs
Rule 3 :	$\dfrac{i : \ldots, \alpha, \alpha \rightarrow \beta}{i+1 : \ldots, \beta}$	Modus ponens
Rule 4 :	$\dfrac{i : \ldots, P_1 a, \ldots, P_n a, (\forall x)[(P_1 x \wedge \ldots \wedge P_n x) \rightarrow Qx]}{i+1 : \ldots, Qa}$	Another version of modus ponens
Rule 5 :	$\dfrac{i : \ldots}{i+1 : \ldots, \neg K(i, \beta)}$	Negative introspection[a]
Rule 6 :	$\dfrac{i : \ldots, \alpha, \neg\alpha}{i+1 : \ldots, Contra(i, \alpha, \neg\alpha)}$	Presence of (direct) contradiction
Rule 7 :	$\dfrac{i : \ldots, \alpha}{i+1 : \ldots, \alpha}$	Inheritance[b]

[a] where β is not a theorem at step i, but is a closed sub-formula at step i.

[b] where nothing of the form $Contra(i-1, \alpha, \beta)$ nor $Contra(i-1, \beta, \alpha)$ is an i-theorem, and where α is not of the form $Now(\beta)$. That is, contradictions and time are not inherited.

The intuitive reason time is not inherited is that time changes at each step. (Clearly, in general one would want a stronger restriction on the inheritance of time. It is not obvious, however, what that should be. This problem is related to the dangling-time-parameter issue discussed on page 15. For the purposes of illustrating certain behaviors, however, a stronger restriction is not necessary.)

The intuitive reason contradicting wffs α and β are not inherited is that not both can be true, and so the agent should, for that reason, be unwilling to simply assume either to be the case without further justification. This does not mean, however, that neither will appear at the next step, for either or both may appear for other reasons, as will be seen. Note also that the wff $Contra(i, \alpha, \neg\alpha)$ *will* be inherited, since it is not itself either time or a contradiction, and (intuitively) it expresses a fact (that there was a contradiction at step i) that remains true.

Figure 4: Rules of inference corresponding to INF_B

might have remained indefinitely. This behavior we regard as within the spirit of the reasoning we wish to study, since it follows real-time vagaries of what is actually done rather than an externally proscribed notion of validity.

Definition .19 *A wff is said to be* P-free *if it does not contain the predicate letter* P.

Definition .20 *An observation-function OBS is said to be* P-free *if* $\forall i \forall \alpha(\alpha \in OBS(i) \rightarrow \alpha$ *is P-free*).

Definition .21 *An observation-function OBS is said to be* valid *if* $\forall i \forall \alpha(\alpha \in OBS(i) \rightarrow \alpha$ *is logically valid*).

Theorem .22 $SL_7(OBS, INF_B)$ *is step-wise consistent if OBS is both valid and Now-free.*

Proof: See the appendix for the details. The idea is to show $SL_7(OBS, INF_B)$ has a step-model, and apply Theorem .16. □

Remark .23 *When OBS is empty (i.e.* $\forall i, OBS(i) = \emptyset$*),* $SL_7(OBS, INF_B)$ *reduces to a ''clock'', i.e.,* $\forall i, SL_7(OBS, INF_B) \vdash_i \alpha$ *iff* $\alpha = Now(i)$.

6.2 The Brother Problem

In this section we use Moore's *Brother problem* (see [30]) to provide examples of $SL_7(\cdot, INF_B)$ at work. One reasons, ''Since I don't know I have a brother, I must not.'' This problem can be broken down into two: the first requires that the reasoner be able to decide he doesn't know he has a brother; the second that, on that basis, he, in fact, does not have a brother (from *modus ponens* and the assumption that ''If I had a brother, I'd know it.'') The first of these seems to lend itself readily to step-logic, in that the negative reflection problem (determining when something is not known) reduces to a simple look-up. As an illustration of how SL_7 works, we present a real-time solution to Moore's *Brother problem*.[25]

In the following three sub-sections, 6.2.1--6.2.3, we present synopses of computer-generated results for three different scenarios where the agent determines whether or not a brother exists. Let B be a 0-argument predicate letter representing the proposition that a brother exists. Let P be a 0-argument predicate letter (other than B) that represents a proposition that implies that a brother exists.[26] In each case, at some step i the agent has the axiom $P \rightarrow B$, and also the following autoepistemic axiom which represents the belief that not knowing B ''now'' implies $\neg B$.

Axiom 1 $(\forall x)[(Now(x) \wedge \neg K(x-1, B)) \rightarrow \neg B]$[27]

The following behaviors are illustrated:

- If B is among the wffs of which the agent is aware at step i, but not one that is believed at step i, then the agent will come to know this fact ($\neg K(i, B)$, that it was not believed at step i) at step $i+1$. As a consequence of this, other information may be deduced. In this case, the agent concludes $\neg B$ from the autoepistemic axiom (Axiom 1). Clearly the Now predicate plays a critical role. Section 6.2.1 below illustrates this case.
- The agent must refrain from such negative introspection when in fact B is already known; see Section 6.2.2.
- A conflict may occur if something is coming to be known while negative introspection is simultaneously leading to its negation. The third illustration (see Section 6.2.3 below) shows this being resolved in an intuitive manner (though not one that will generalize as much as we would like; this is an area we are currently exploring).

[25] We use this problem although, according to Moore, it technically does not involve ''true'' default reasoning. We could as easily have used a standard simple default such as one about birds typically flying. Also, note that there is a wealth of background commonsense knowledge not usually made explicit in formal treatments, such as that brothers, if they exist, are known not merely at one moment but by repeated experience over long periods of time. However, we will not attempt a detailed formal treatment of such fine points, as they belong to a different domain: naive perceptual psychophysics.

[26] P might be something like ''My parents have two sons,'' together with appropriate axioms.

[27] It appears that some arithmetic is involved here, but simple syntactic devices can obviate any genuine subtraction. We can replace, for instance, $K(i-1, \alpha)$ by $J(i, \alpha)$ with the intuitive meaning that α was known ''just a moment ago'', i.e., at i. Alternatively, we can use successor notation for natural numbers.

6.2.1 Simple negative introspection succeeds

In this example the agent is not able to deduce the proposition B, that he has a brother, and hence is able to deduce $\neg B$, that he does *not* have a brother. See Figure 5. Here, and in Sections 6.2.2 and 6.2.3, for ease of reading we underline in each step those wffs which are new (i.e., which appear through other than inheritance). For the purposes of illustration, let i be arbitrary and let

$$OBS_{B_1}(j) = \begin{cases} \{P \rightarrow B, (\forall x)[(Now(x) \wedge \neg K(x-1, B)) \rightarrow \neg B]\} & \text{if } j = i \\ \emptyset & \text{otherwise} \end{cases}$$

Since B is not an i-observation (and thus is not an i-theorem), the agent uses Rule 5, the negative introspection rule, to conclude $\neg K(i, B)$ at step $i + 1$. At step $i + 2$ the agent concludes $\neg B$ from the autoepistemic knowledge stated above (Axiom 1) and the use of the alternate version of modus ponens, Rule 4.

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

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

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

Figure 5: Negative introspection succeeds

6.2.2 Simple negative introspection fails (appropriately)

In this example, let

$$OBS_{B_2}(j) = \begin{cases} \{P \rightarrow B, (\forall x)[(Now(x) \wedge \neg K(x-1, B)) \rightarrow \neg B], B\} & \text{if } j = i \\ \emptyset & \text{otherwise} \end{cases}$$

Thus the agent has B at step i, and is blocked (appropriately for this example) from deducing at step $i + 1$ the wffs $\neg K(i, B)$ and $\neg B$. See Figure 6.

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

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

Figure 6: Negative introspection fails appropriately

Note that a traditional final-tray-like approach could produce quite similar behavior to that seen in Figures 5 and 6 if it is endowed with a suitable introspection device, although it would not have the real-time step-like character we are trying to achieve.

6.2.3 Introspection contradicts other deduction

In this example, let

$$OBS_{B_3}(j) = \begin{cases} \{P \rightarrow B, (\forall x)[(Now(x) \wedge \neg K(x-1, B)) \rightarrow \neg B], P\} & \text{if } j = i \\ \emptyset & \text{otherwise} \end{cases}$$

In Figure 7 we see then that the agent does not have B at step i, but is able to *deduce* B at step $i+1$ from $P \rightarrow B$ and P at step i. Since the agent is *aware* (in our sense) of B at step i, and yet does not have B as a *conclusion* at i, it will deduce $\neg K(i, B)$ at step $i+1$. Thus both B and $\neg K(i, B)$ are concluded at step $i+1$. At step $i+2$ Axiom 1 (the autoepistemic axiom), together with $Now(i+1)$ and $\neg K(i, B)$ and Rule 4, will produce $\neg B$. A conflict results, which is noted at step $i+3$. This then inhibits inheritance of both B and $\neg B$ at step $i+4$. Although neither B nor $\neg B$ is *inherited* to step $i+4$, B is *re-deduced* at step $i+4$ via modus ponens from step $i+3$. Thus B ''wins out'' over $\neg B$ due to its existing justification in other wffs, while $\neg B$'s justification is ''too old'': $\neg K(i+2, B)$, rather than $\neg K(i, B)$, would be needed. We see then that the conflict resolves due to the special nature of the time-bound ''now'' feature of introspection.

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

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

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

$i+3$: $\underline{Now(i+3)}, P \rightarrow B, (\forall x)[(Now(x) \wedge \neg K(x-1, B)) \rightarrow \neg B], P, B, \neg K(i, B), \neg K(i, \neg B),$
$\neg B, \neg K(i+1, \neg B), \underline{Contra(i+2, B, \neg B)}$

$i+4$: $\underline{Now(i+4)}, P \rightarrow B, (\forall x)[(Now(x) \wedge \neg K(x-1, B)) \rightarrow \neg B], P, \neg K(i, B), \neg K(i, \neg B)$
$\neg K(i+1, \neg B), Contra(i+2, B, \neg B), \underline{B}, \underline{Contra(i+3, B, \neg B)}$

Figure 7: Introspection conflicts with other deduction and resolves

A traditional final-tray-like approach would encounter difficulties with this third example, for at step $i+2$ there is a contradiction. This means that the final tray for a tray-like model of a reasoning agent would simply be filled with all wffs in the language---and no basis for a resolution is possible *within* such a logic.

Remark .24 *The following are true about the consistency of each of the SL_7 theories given in the brother examples:*

1. $SL_7(OBS_{B_1}, INF_B)$ *is step-wise consistent.*
2. $SL_7(OBS_{B_2}, INF_B)$ *is step-wise consistent.*
3. $SL_7(OBS_{B_3}, INF_B)$ *is eventually consistent (but not step-wise consistent*[28]*).*

Proof: We briefly sketch the proof of part 1 of the preceding remark. Parts 2 and 3 are similar; part 3 involves constructing a model for each step after the last inconsistent step (which happens to be step $i+3$).

Since $OBS_{B_1}(j) = \emptyset$, for $j < i$, by Remark .23, if $j < i$, $\alpha \in \vdash_j$ iff $\alpha = Now(j)$. Therefore every step in $SL_7(OBS_{B_1}, INF_B)$ up to and including step $i-1$ is consistent. From step i on we have additional theorems which must be considered. This is due to the fact that $OBS_{B_1}(i)$ is not empty. To show that step i and all subsequent steps are consistent, we propose a model M_j for each step j. In each M_j interpret the predicates in the following way: $K \equiv false, B \equiv false, P \equiv false, Now(k) \equiv k = j$, where P is any predicate other than K, B, or Now.

[28] This is why a traditional final-tray-like approach would encounter difficulties with this example.

We can then see that we have a model for each of steps i thru $i+2$. Noting that for an arbitrary step $i+k, k > 2$,

$$\vdash_{i+k} = \left\{ \begin{array}{l} Now(i+k), \\ P \rightarrow B, \\ (\forall x)[(Now(x) \wedge \neg K(x-1, B)) \rightarrow \neg B], \\ \neg B, \\ \neg K(i, \neg B), \neg K(i+1, \neg B), \\ \neg K(i, B), \ldots, \neg K(i+k-1, B), \\ \neg K(i, P), \ldots, \neg K(i+k-1, P) \end{array} \right\}$$

we see that, again, M_{i+k} is an appropriate model. Therefore, by Theorem .16, $SL_7(OBS_{B_1}, INF_B)$ is step-wise consistent. □

7 Discussion and Future Work

We have argued that explicit representation of individual reasoning steps *as they occur* will prove crucial for many problems in the formalization of commonsense reasoning. We are developing step-logic for this purpose. Specifically, we have implemented a real-time inference mechanism that correctly infers a lack of knowledge of certain wffs; that correctly will not infer a lack when the knowledge is in fact present; and that correctly resolves a contradiction when timing is such that new knowledge arises conflicting with a prior (or simultaneous) conclusion of its lack. Of course, we have shown this only in a very limited context.

One of the difficulties we encountered in our efforts to represent real-time reasoning involved the concept of ''now.'' The approach we have found most useful so far is the one given in Rules 1 and 7, coupled with Rules 3 and 4 for modus ponens (see Figure 4), where Now is a predicate symbol with special treatment regarding inheritance. However, there are variations on our example where this is not completely satisfactory. In particular, a difficulty can arise when there is a detachment of a ''$Now(j)$'' sub-formula from the rest of the formula, producing what we call a ''dangling time parameter.''

For example, if in Figure 7, instead of Axiom 1, we had used the following:

Axiom 2 $(\forall x)[Now(x) \rightarrow (\neg K(x-1, B) \rightarrow \neg B)]$,

then an intermediate conclusion, namely,

(1) $\neg K(i, B) \rightarrow \neg B$,

would have occurred at step $i+2$. The problem is that (1) inherits to future steps, even though the intended significance of i in (1) was that it was the *current* time step (i.e., linked to $Now(i)$) rather than any particular fixed step; by step $i+2$, the term i has lost its tie to the wff $Now(i)$, and so ''dangles'' inappropriately. Modus ponens can then be used with (1) to conclude $\neg B$ at any step after $i+2$ in which $\neg K(i, B)$ appears. Since we do have $\neg K(i, B)$ at step $i+1$ and all subsequent steps (inherited via Rule 7), the conclusion $\neg B$ is re-deduced from step $i+3$ on, despite the contradiction resolution mechanism we have discussed.

This emphasizes that B's merely not being known some time ago is insufficient reason to conclude $\neg B$. That is, if we have deduced $\neg B$ from $\neg K(i, B)$ at step $i+2$, but later (or in the meantime) we conclude B, we no longer want to be able to deduce $\neg B$. Any satisfactory treatment, then, should refer to the fact that the agent does not know B at the *current* time step, before autoepistemically deducing $\neg B$. The particular formulation of the *Brother problem* that we presented in Section 6.2 satisfies this condition due to the special form of the autoepistemic axiom (Axiom 1). A similar, but even safer, approach is that of employing a special purpose *inference rule* (instead of the autoepistemic *axiom*) such as:

(2) $$\frac{Now(i) \wedge \neg K(i-1, B)}{\neg B}$$

However we prefer a treatment that allows the agent to explicitly represent such a train of deduction, as in Axiom 1, for then the agent also has the possibility of reasoning about this very process of reasoning. On the other hand, the fact

that the alternate version (Axiom 2) above is not satisfactory suggests that dangling time parameters be avoided in a more general (less syntax-dependent) way. We are currently working on this.

In addition to further exploration of the dangling-time-parameter problem, further mechanisms are desirable for handling contradictions. For instance, if B *and* $\neg B$ are deducible from other beliefs $P_1, \ldots, P_n$ (without the use of the introspective rule, so that earlier steps contain indirect contradictions), it is not enough to block inheritance of B and $\neg B$. Rather the roots of the contradiction, $P_1, \ldots, P_n$, must be investigated in order to unwind the contradiction. One interesting feature, however, is that it is not at all critical whether a contradiction is *instantly* resolved. The swamping problem is much less serious than in final-tray-like logics. In step-logic, the agent can continue reasoning step-by-step in the presence of a contradiction. The possible ''spread'' of invalid conclusions from a contradiction itself goes only step-by-step, thus presenting the possibility of controlling it by effective means.

However, it should be emphasized that ''resolving'' a contradiction, as opposed simply to preventing it from upsetting ones' reasoning, is not necessarily a good thing. Consider again the *Nell and Dudley problem.* If Dudley's reasoning uncovers a contradiction it may be best that he ignore it---that is, that he see it and put it aside, rather than stop to examine its sources in an effort to resolve it as the train rushes toward Nell. Thus the spirit of step-logics is toward unencumbered and effective reasoning more than the traditional stricture of logical consistency.

Another approach to contradictions is as follows: in addition to stopping the inheritance of contradictands, we can also disallow their use as antecedents to certain inference rules. We intend to explore this approach in future work.

On the other hand, it appears that any step-logic appropriate for broad commonsense reasoning should be self-stabilizing, rather than continually faced with emerging or inherited contradictions---unless of course the environment is unfriendly enough to maintain an unlimited (infinite) supply of new contradictions to current beliefs. We regard this as an acid test of the long-range significance of our approach, which we formulate as a conjecture:

Conjecture .25 *There is a powerful and natural class of self-stabilizing step-logics.*

Although the present paper makes only very modest use of a retraction mechanism, we expect retraction to play a much greater role in more sophisticated versions of SL_7. We intend eventually to make broad use of such a mechanism in order to keep the belief set at any given step to a reasonable size. We anticipate the introduction of a notion of relevance, where the beliefs that are ''relevant'' to the current situation are the only ones in the current belief set. This will also lend itself to the re-inclusion of tautologies and other logical axioms. In sequel papers, we will address this relevance problem as well as the further aspects of inconsistency mentioned above, the dangling-time-parameter problem, step-models for inconsistent theories, and a real-time solution to the *Three-wise-men problem*.

In this paper we have argued that a formal treatment of commonsense reasoning situated in time is not only possible but can remain largely deductive in character. We have indicated certain key points of difference from more traditional deductive mechanisms. In particular, negative introspection becomes computationally tractable, while also forcing a ''time-delay'' between knowledge and self-knowledge. Another difference is tolerance for contradictions, found in step-logic but not in traditional logic. We illustrated this with some rather simple examples of default reasoning. Finally, we have outlined several areas for further study, particularly a more ambitious retraction mechanism suitable both for the problem of relevance and for deeper probing of contradictions.

8 Acknowledgments

We would like to thank Bill Gasarch, Jorge Lobo, Michael Miller, Jack Minker, Irene Durand, and several referees for helpful observations and suggestions.

A Proof of Theorem .22

Proof: We show $SL_7(OBS, INF_B)$ has a step-model, and apply Theorem .16. Let M_i be such that:

1. (MODEL-NOW) $M_i \models Now(x)$ iff $x = i$.
2. (MODEL-K) $M_i \models K(j, \alpha)$ iff $\vdash_j \alpha$.

3. (MODEL-P) $M_i \not\models P(x_1, \ldots, x_n)$ if P is a predicate symbol other than Now or K.

We show $M =< M_0, M_1, \ldots, M_i, \ldots >$ a step-model for $SL_7(OBS, INF_B)$ by induction on the index.
For each index i, we want to show the following:[29]

1. (HYP.CONTRA) $Contra(\alpha, \beta, \gamma) \notin \vdash_i$.
2. (HYP.NOW) If $M_i \models K(i, \alpha)$ and α is not Now-free, then $\alpha = Now(i)$.
3. (HYP.MODEL) $M_i \models \alpha$ if $\vdash_i \alpha$.
4. (HYP.CONSISTENT) $\vdash_i$ is consistent.

Base case: $i = 0$

1. (HYP.CONTRA) This is true since $\vdash_0$ is empty.
2. (HYP.NOW) By (MODEL-K), $M_0 \models K(0, \alpha)$ iff $\vdash_0 \alpha$.
   Since $\vdash_0$ is empty, $M_0 \not\models K(0, \alpha)$ for any α.
   Therefore, this hypothesis is trivially true.
3. (HYP.MODEL) Since $\vdash_0$ is empty, this hypothesis is trivially true.
4. (HYP.CONSISTENT) $\vdash_0$ is consistent since it is empty.

Assume Hypotheses (HYP.CONTRA), (HYP.NOW), (HYP.MODEL), and (HYP.CONSISTENT) for $i - 1$. We must show these are true for i.

1. (HYP.CONTRA) To show $Contra(\alpha, \beta, \gamma) \notin \vdash_i$.
   By INF_B, $Contra(\alpha, \beta, \gamma) \in \vdash_i$ only thru the Rules 1-7. But:
   (a) Rule 1 will not bring in any wffs of the form $Contra(\alpha, \beta, \gamma)$.
   (b) Rule 2 will not bring in any wffs of the form $Contra(\alpha, \beta, \gamma)$.
   (c) Suppose $\delta, \delta \rightarrow Contra(\alpha, \beta, \gamma) \in \vdash_{i-1}$.
   Then, by Hyp. (HYP.MODEL), $M_{i-1} \models \delta$ and
   $M_{i-1} \models \delta \rightarrow Contra(\alpha, \beta, \gamma)$.
   Hence, since M_{i-1} is an interpretation, $M_{i-1} \models Contra(\alpha, \beta, \gamma)$.
   But, by (MODEL-P), $M_{i-1} \not\models Contra(\alpha, \beta, \gamma)$. $\longrightarrow\!\!\leftarrow$
   Thus both δ and $\delta \rightarrow Contra(\alpha, \beta, \gamma)$ cannot be $\in \vdash_{i-1}$.
   Therefore Rule 3 will not produce $Contra(\alpha, \beta, \gamma)$ at step i.
   (d) Suppose $P_1a, \ldots, P_na, \forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Contra(\alpha, \beta, \gamma)] \in \vdash_{i-1}$.
   Then, by Hyp. (HYP.MODEL), $M_{i-1} \models P_1a$ and, $\ldots$, and $M_{i-1} \models P_na$ and
   $M_{i-1} \models \forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Contra(\alpha, \beta, \gamma)]$.
   Hence, since M_{i-1} is an interpretation, $M_{i-1} \models Contra(\alpha, \beta, \gamma)$.
   But, by (MODEL-P), $M_{i-1} \not\models Contra(\alpha, \beta, \gamma)$. $\longrightarrow\!\!\leftarrow$
   Thus $P_1a, \ldots, P_na, \forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Contra(\alpha, \beta, \gamma)]$ cannot all be $\in \vdash_{i-1}$.
   Therefore Rule 4 will not produce $Contra(\alpha, \beta, \gamma)$ at step i.
   (e) Rule 5 will not bring in any wffs of the form $Contra(\alpha, \beta, \gamma)$.
   (f) By inductive hyp. (HYP.CONSISTENT), $\vdash_{i-1}$ is consistent.
   Thus α and $\neg\alpha$ cannot both be $\in \vdash_{i-1}$.
   Therefore, Rule 6 will not apply.
   (g) By the inductive hypothesis, $Contra(\alpha, \beta, \gamma) \notin \vdash_{i-1}$.
   Therefore Rule 7 will not bring in any wffs of the form $Contra(\alpha, \beta, \gamma)$.

   Therefore $Contra(\alpha, \beta, \gamma) \notin \vdash_i$.

[29] A step-model requires that $M_i \models K(j, \alpha)$ iff $\vdash_j \alpha$. This we know to be true $\forall i$ and $\forall j$ directly from (MODEL-K).

2. (HYP.NOW) Suppose $M_i \models K(i, \alpha)$ and α is not Now-free. To show $\alpha = Now(i)$.
   From (MODEL-K), $\vdash_i \alpha$.
   By INF_B, either:
   (a) $\alpha = Now(i)$.
   (b) $\alpha \in OBS(i-1)$. Since OBS is Now-free, Now doesn't appear in α. $\longrightarrow\!\!\leftarrow$
   (c) $\beta, \beta \rightarrow \alpha \in \vdash_{i-1}$.
   Then, by (MODEL-K), $M_{i-1} \models K(i-1, \beta \rightarrow \alpha)$.
   Since α is not Now-free, $\beta \rightarrow \alpha$ is not Now-free.
   But by the inductive hypothesis, if $\beta \rightarrow \alpha$ is not Now-free, then $\beta \rightarrow \alpha$ is $Now(i-1)$. $\longrightarrow\!\!\leftarrow$
   (d) $\alpha = Qa$ and $P_1a, \ldots, P_na, \forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Q(x)] \in \vdash_{i-1}$.
   Since α contains Now, Q must be Now.
   Then by (MODEL-K), $M_{i-1} \models K(i-1, \forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Now(x)])$.
   But, by the inductive hypothesis, $\forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Now(x)])$ must be $Now(i-1)$. $\longrightarrow\!\!\leftarrow$
   (e) $\alpha = \neg K(i-1, \beta)$ and $\beta \notin \vdash_{i-1}$ and $\gamma \in \vdash_{i-1}$, where β is a closed sub-formula of γ.
   Since Now appears in α, Now must also appear in β, and thus must also appear in γ.
   Now, by (MODEL-K), $M_{i-1} \models K(i-1, \gamma)$.
   But then, by the inductive hypothesis, $\gamma = Now(i-1)$. Hence $\beta = \gamma$.
   But then, $\beta \in \vdash_{i-1}$. $\longrightarrow\!\!\leftarrow$
   (f) $\alpha = Contra(i-1, \beta, \neg\beta)$.
   But by (HYP.CONTRA), $Contra(\alpha, \beta, \gamma) \notin \vdash_i$. $\longrightarrow\!\!\leftarrow$
   (g) $\alpha \in \vdash_{i-1}$ and $\alpha \neq Now(\beta)$ and $Contra(i-2, \alpha, \gamma) \notin \vdash_{i-1}$ and $Contra(i-2, \gamma, \alpha) \notin \vdash_{i-1}$.
   Then, by (MODEL-K), $M_{i-1} \models K(i-1, \alpha)$.
   Then by the inductive hypothesis, $\alpha = Now(i-1)$. $\longrightarrow\!\!\leftarrow$
   Therefore, if $M_i \models K(i, \alpha)$ and α is not Now-free, then $\alpha = Now(i)$.
3. (HYP.MODEL) Let $\alpha \in \vdash_i$. To show $M_i \models \alpha$.
   By INF_B, either:
   (a) $\alpha = Now(i)$.
   Then by (MODEL-NOW), $M_i \models \alpha$.
   (b) $\alpha \in OBS(i-1)$. Then α is valid.
   Hence α is true in any interpretation; and in particular, $M_i \models \alpha$.
   (c) $\beta, \beta \rightarrow \alpha \in \vdash_{i-1}$.
   Then, by the inductive hypothesis, $M_{i-1} \models \beta$ and $M_{i-1} \models \beta \rightarrow \alpha$.
   Hence, since M_{i-1} is an interpretation, $M_{i-1} \models \alpha$.
   Now, by (MODEL-K), $M_{i-1} \models K(i-1, \beta \rightarrow \alpha)$.
   And by Hyp. (HYP.NOW), if Now appears in $\beta \rightarrow \alpha$, then $\beta \rightarrow \alpha$ is $Now(i-1)$. $\longrightarrow\!\!\leftarrow$
   Therefore, $\beta \rightarrow \alpha$ is Now-free; hence α is Now-free.
   Then by Lemma .26, $M_i \models \alpha$.
   (d) $\alpha = Qa$ and $P_1a, \ldots, P_na, \forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Q(x)] \in \vdash_{i-1}$.
   Then, by the inductive hypothesis,
   $M_{i-1} \models P_1a$ and ... and $M_{i-1} \models P_na$ and $M_{i-1} \models \forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Q(x)]$.
   Hence, since M_{i-1} is an interpretation, $M_{i-1} \models Qa$, i.e. $M_{i-1} \models \alpha$.
   Now, by (MODEL-K), $M_{i-1} \models K(i-1, \forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Q(x)])$.
   And by Hyp. (HYP.NOW), if Now appears in $\forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Q(x)]$,
   then $\forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Q(x)]$ is $Now(i-1)$. $\longrightarrow\!\!\leftarrow$
   Therefore, $\forall x[(P_1x \wedge \ldots \wedge P_nx) \rightarrow Q(x)]$ is Now-free; and in particular, Q is not Now.
   Hence α is Now-free.
   Then by Lemma .26, $M_i \models \alpha$.

(e) $\alpha = \neg K(i-1, \beta)$ and $\beta \notin \vdash_{i-1}$ and $\gamma \in \vdash_{i-1}$, where β is a closed sub-formula of γ.
By (MODEL-K), $M_{i-1} \not\models K(i-1, \beta)$.
And since M_{i-1} is an interpretation, $M_{i-1} \models \neg K(i-1, \beta)$, i.e. $M_{i-1} \models \alpha$.
To show α is Now-free, it is sufficient to show β is Now-free.
Now, by (MODEL-K), $M_{i-1} \models K(i-1, \gamma)$.
And by Hyp. (HYP.NOW), if Now appears in γ, then $\gamma = Now(i-1)$.
Now β is a closed sub-formula of γ, hence $\beta = Now(i-1)$. Then $\beta \in \vdash_{i-1}$. $\longrightarrow\!\leftarrow$
Therefore γ is Now-free; hence, β is Now-free.
Since β is Now-free, α is also Now-free.
Then by Lemma .26, $M_i \models \alpha$.

(f) $\alpha = Contra(i-1, \beta, \neg\beta)$ and $\beta, \neg\beta \in \vdash_{i-1}$.
But by Hyp. (HYP.CONSISTENT), we cannot have both $\beta \in \vdash_{i-1}$ and $\neg\beta \in \vdash_{i-1}$.
Therefore $\alpha \neq Contra(i-1, \beta, \neg\beta)$. $\longrightarrow\!\leftarrow$

(g) $\alpha \in \vdash_{i-1}$ and $\alpha \neq Now(\beta)$ and $Contra(i-2, \alpha, \gamma) \notin \vdash_{i-1}$ and $Contra(i-2, \gamma, \alpha) \notin \vdash_{i-1}$.
Now, by (MODEL-K), $M_{i-1} \models K(i-1, \alpha)$.
And by Hyp. (HYP.NOW), if Now appears in α, $\alpha = Now(i-1)$. $\longrightarrow\!\leftarrow$
Therefore, α is Now-free.
Now by the inductive hypothesis, $M_{i-1} \models \alpha$.
Then by Lemma .26, $M_i \models \alpha$.

Therefore, if $\alpha \in \vdash_i$, then $M_i \models \alpha$.

4. (HYP.CONSISTENT) To show $\vdash_i$ is consistent.
Suppose $\vdash_i$ is inconsistent. Then there exist wffs $\alpha_1, \ldots, \alpha_n \in \vdash_i$ which are mutually inconsistent.
By Hyp. (HYP.MODEL), $M_i \models \alpha_1$ and ... and $M_i \models \alpha_n$.
But since M_i is an interpretation, $\alpha_1, \ldots, \alpha_n$ cannot be mutually inconsistent.
Therefore, $\vdash_i$ is consistent.

Therefore, by induction we have shown that (HYP.CONTRA), (HYP.NOW), (HYP.MODEL), and (HYP.CONSISTENT) hold for all i.

Now, (HYP.MODEL) shows that $M =< M_0, M_1, \ldots, M_i, \ldots >$ is a step-model for $SL_7(OBS, INF_B)$.

And by Theorem .16, $SL_7(OBS, INF_B)$ is step-wise consistent.

(We also have step-wise consistency directly from (HYP.CONSISTENT).) □

Lemma .26 *$M_i \models \alpha$ if $M_{i-1} \models \alpha$ and α is Now-free.*

Proof: By (MODEL-K), if a wff $K(j, \beta)$ is true in some M_i, then it is true in every M_i.

Likewise, if a wff $K(j, \beta)$ is false in some M_i, then it is false in every M_i.

By (MODEL-P), wffs $P(x_1, \ldots, x_n)$, where the predicate symbol P is neither Now nor K, are false in every M_i.

It follows that every Now-free wff α will either be true in every M_i or false in every M_i, for such wffs will be built out of wffs $K(j, \beta)$ and $P(t_1, \ldots, t_n)$ whose truth-values do not change with i.

Therefore, if $M_{i-1} \models \alpha$ and α is Now-free, then $M_i \models \alpha$. □

References

[1] 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.

[2] 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.

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

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

[5] J. Goodwin. *A Theory and System for Non-Monotonic Reasoning*. PhD thesis, Department of Computer and Information Science, Linkoping University, Linkoping, Sweden, 1987.

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

[7] D. Perlis. Non-monotonicity and real-time reasoning. In *Proceedings of the Workshop on Non-monotonic Reasoning*. AAAI, October 1984. New Paltz, New York.

[8] 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.

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

[10] K. Konolige. A deductive model of belief. In *Proceedings of the 8th Int'l Joint Conf. on Artificial Intelligence*, pages 377--381, 1983. Karlsruhe, West Germany.

[11] H. Levesque. A logic of implicit and explicit belief. In *Proceedings of the 3rd National Conf. on Artificial Intelligence*, pages 198--202, 1984. Austin, TX.

[12] R. Fagin and Y. Halpern, J. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39--76, 1988.

[13] 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.

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

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

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

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

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

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

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

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

[22] E. McKenzie and R. Snodgrass. Extending the relational algebra to support transaction time. In *Proceedings of the SIGMOD Conference*, pages 454--466. ACM, 1987. San Francisco, California.

[23] E. Gettier. Is justified true belief knowledge? *Analysis*, 23:121--123, 1963.

[24] J. Drapkin and D. Perlis. Analytic completeness in sl_0. Technical Report TR-1682, Department of Computer Science, University of Maryland, College Park, Maryland, 1986.

[25] E. Mendelson. *Introduction to Mathematical Logic*. Wadsworth, Belmont, CA, 3rd edition, 1987.

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

[27] R. Weyhrauch. The building of mind. In *1986 Workshop on Meta-Architectures and Reflection*, 1986. Sardinia, Italy.

[28] J. Doyle. Some theories of reasoned assumptions: An essay in rational psychology. Technical report, Department of Computer Science, Carnegie Mellon University, 1982.

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

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