The Internal Reasoning of Robots

Don Perlis, Justin Brody, Sarit Kraus, Michael Miller

University of Maryland, Goucher College, Bar Ilan University, Bethesda MD

perlis@cs.umd.edu, justin.brody@goucher.edu, sarit@cs.biu.ac.il, mjmiller@gmail.com,

Abstract

We argue for the value of examining the internal processes that robots might actually use to draw inferences in a timely way in a dynamic world. This requires a significantly different way of thinking about logic and reasoning, which in turn bears on some traditional logic-related problems such as omniscience and reasoning in the presence of a contradiction, as well as on a wide variety of other AI issues. A non-standard internally-evolving notion of time seems to be the key that unlocks other tools.

Introduction

We teeter on the edge of the age of general-purpose robots. It thus becomes ever more important that commonsense reasoning (CSR) examine in some detail just how such a robot will actually think, i.e., produce inferences over time (as it plans, decides, assesses, questions, learns, explores, updates, reconsiders, etc). In particular, robots will need to keep their reasoning abreast of at least some aspects of the evolving world, including the passage of time and how they are progressing with regard to their own (also evolving) goals.[1]

On the surface much of CSR may seem to be aiming at just these issues.[2] But the bulk of such work follows what Ray Reiter has called the "external design stance" (Reiter 2001, pp 292-293): that of a designer-scientist "entirely external to … [and] … looking down on some world inhabited by an agent." Indeed, a lot of this work is very relevant and has led to major advances in our understanding: situation calculus, nonmonotonic reasoning, and much more. Still, the external stance is nevertheless a very highly idealized abstraction that creates an unworkable barrier regarding a robot's internal reasoning, and in addition faces huge hurdles such as omniscience, contradiction-intolerance, and more.

---

[1] While we recognize that Markov decision processes (MDPs) and related technical tools are standard items in much of current (often highly-structured special-task) robotic work, general-purpose robots will be bombarded with "culturally supplied" information from other agents, signage, online, and so on, and will need to reason in real-time with such information. Hence a knowledge base (KB) managed in large measure by inferential processes seems unavoidable.

[2] See for instance (Rajan&Saffiotti 2017) for very recent work.

This paper attempts to shed light on that barrier and those hurdles, and to highlight an alternative that drives a sharp wedge between two notions of logic: (i) the standard "external" kind (*E-logics*) that specify features from afar via closure under (some form of) consequence or entailment relation, and (ii) "internal" ones (*I-logics*) that represent (and indeed can actually be used for) the inferential processing undertaken by an agent over time. (We especially focus on *active logic*, which is perhaps the most developed form of I-logic so far. Active logic grew out of ideas in (Elgot-Drapkin&Perlis 1990), and has been continually investigated ever since (Nirkhe et al 1991; Miller&Perlis 1996; Kraus et al 2000; Anderson et al 2008; Brody et al 2014; Brody&Perlis 2015).)

As we will see, some of the issues faced by E-logics (e.g., omniscience) simply go away in an I-logic approach. In addition, we have found a wide array of unexpected benefits of such an approach, that ties CSR to many other parts of AI. Thus the present paper is also a kind of progress report, pulling together many aspects of our attempt to look under the robotic hood, to craft appropriate logic mechanisms to go there, and to explore applications across AI. As such, it will have a large number of short sections; we beg the reader's indulgence, for we see this as the most useful way to communicate the range of these ideas compactly.

The single most salient departure that I-logics make from E-logics is that of *taking into account the actual process of inferring as something that itself takes time*. Thus when a conclusion is inferred, it has become a later time than prior to reaching that conclusion. This time-stratification spreads successive inferences out and leaves a self-updating record of an agent's evolving beliefs up until the present moment (which itself then moves ahead one more step, and so on indefinitely). Secondly, this stratification then provides a very simple yet *far-reaching form of introspection*: looking back at one's beliefs of past moments and drawing conclusions bearing on everything from non-monotonicity and contradiction-handling, to ambiguity resolution, agent control of semantics, and awareness of own actions. Third, the notions of axiom and theorem and entailment are no longer very informative: beliefs come and go – still due to (various forms of) inference, but including evolving time and

the ability to give up (i,e., disinherit) beliefs that are judged as no longer appropriate.

Active logic in particular posits an unending[3] sequence of time-steps, at each of which the knowledge base (KB) has a finite number of wffs, considered as the beliefs that the reasoning agent holds (at that step); the contents of the KB then fluctuate in time, and there is no final state where the agent arrives at its "finished" belief-set. It is the agent's behavior through time that is of interest.

Elementary Example: Go to Lunch

A robot needs to get to a noon lunch date, and it is now 11am. How can it ever decide to start walking? The problem is that, given Now(11:00), standard logics will treat this as an axiom and so the robot will never realize the time has changed, e.g., that it has now become 11:30 and it should start walking.[4] Clearly it is essential that the robot be able to update its belief as to what time it is.

An example of the desired behavior is illustrated below; underlined items on each line indicate beliefs newly-formed at the corresponding time-step:

Time	Evolving belief set
11:00	Now(11:00); Now(11:30) → Do(walk)
11:01	Now(11:01); Now(11:30) → Do(walk)
…	
11:30	Now(11:30); Now(11:30) → Do(walk)
11:31	Now(11:31); Now(11:30)→Do(walk), Do(walk)

At time 11:31 it has just inferred Do(walk).[5] Notice that beliefs of the form Now(t) come and go, whereas the "plan" to walk starting at 11:30 continues to be inherited.[6] A "clock" inference rule (along with Modus Ponens in the last two steps) can achieve this: from Now(t) infer Now(t+1):

```
t:   Now(t)
-----------------
t+1: Now(t+1)
```

---

[3] In concert with Nilsson's notion of an agent with a lifetime of its own (Nilsson 1983).
[4] If lunch for a robot sounds silly, the reader is invited to imagine that the task instead is to approach and disarm a bomb at noon (when local civilians will have been safely moved away).
[5] If one wants to be picky, perhaps this should have been inferred a little earlier, say at 11:29, so that the walking can actually start by time 11:30. Here we are ignoring such details, and also the granularity of time steps.
[6] After 11:30, there is no need to continue inheriting the plan; current implementations of active logic do not take advantage of this "garbage collection" but we expect our next version to do so.

While this may seem simple enough, it radically changes the notion of a logic from an *external specification* (E-logic) of a system in another world, to an *internal mechanism* (I-logic) operating within and as part of that world. In particular, the example is written in the notation of active logic, the I-logic approach that we have been pursuing.

We next offer three clarifications to avoid confusion between E- and I-logics.

This Is Not Your Grandmother's Temporal Logic
Temporal logics are well known.[7] But, in virtually all cases, they are not *properly* temporal – that is, they do not vary with time. In fact, they are examples of E-logics, taking an external timeless stance even while looking in on a world that may evolve in time. In effect, temporal logics have a frozen permanent *now* from which they can express facts about what is, will be, or was the case at various specified moments. But inferences made using such logics do not correspond to anything changing within the world being explored.

Yet a wealth of beneficial connections arise between a *properly* temporal (I-logic) version of CSR and much of the rest of AI – e.g., NLP, perception, robotics, planning. As noted, this paper attempts to bring together a wide range of such benefits as well as provide motivation for the underlying logical apparatus, especially in the active logic form of I-logic. In effect, time-change is the root out of which all the rest flows. In particular, it *dispenses with omniscience quite trivially*: an agent believes only what it has had time to come to believe *so far*; anything else it may come to believe only later on (as further inferences are drawn). Such an agent certainly does not believe (contain in its KB) all wffs that are entailed by its current beliefs. Indeed, current beliefs may well be inconsistent – more on that below.

This Is Not Your Grandfather's Belief Revision
Belief revision[8] provides a possible way to view the above clock rule: insert Now(11:30) as an update, which triggers relaxation of the KB – removal of Now(11:00) among other changes. Yet that last phrase ("among other changes") is where E-logic reveals one of its main hurdles: standard notions of belief revision – being based on a notion of closure under consequence – cannot serve as a mechanism for a robot to use, simply because such closure in general is very expensive (in most cases non-terminating or even undecidable). This is the omniscience problem, and is uni-

---

[7] For standard approaches, see (Pnueli 1977; Baral&Zhao 2008; Gonzalez et al 2002; Barringer et al 2013; Kraus&Lehmann 1986)
[8] See, e.g., (Gardenfors 2003; Sloan&Turan 1999; Goldsmith et al 2004; Delgrande et al 2013; Diller et al 2015) for traditional E-logic approaches.

versally recognized as unrealistic: producing consequences is time-consuming.[9]

Traditional (E-logic) belief revision also suffers from "recency prejudice" (Perlis 1997, 2000), in which newly acquired information is taken to have a firm validity that pre-existing beliefs must yield to. Yet it is hard to think of a case in which a new item P should take precedence over one's entire KB. The reasons for preferring P would surely in large measure be deeply embedded in that very KB as part of one's understanding of many relevant aspects of the world. Thus P and the KB (including information as to where this new P came from) would need to "fight it out" as to whether to accept P or not; and any conclusion could vary over time as the agent devotes more thought to the matter (and/or may decide to seek more information).

Goodbye to Axioms

Very little in CSR can reasonably be taken as firmly given over an agent's lifetime. Perhaps some mathematical concepts, perhaps some definitions. But more commonly, we hold beliefs for awhile and then relax them if sufficient counterevidence arises. Or, in many cases, we already have that evidence, in the form of other beliefs to the effect that something is in flux (the time, an airplane's location, and so on); sometimes change is the rule. It is hard then to find much to take as axiomatic. Here are two more examples.

(1) Your eleven-year-old son tells you that Barack Obama is 6'8" tall. You do not take this as a fact; on the contrary – although you may not have any specific height in mind for Obama – you do believe 6'8" is sufficiently unusual (and presidents are sufficiently in the news) that it would have been remarked on a lot and you would have heard it before. So you discount the information from your son. But if your son then tells you that Obama has been slouching so as to disguise his height ever since his twenties, and that he is in fact 6'8", would you still be so sure he is wrong?

(2) You hear the TV meteorologist say that the temperature dropped to 1 degree below zero last night; and you accept this. But you would not be especially startled to learn later that the meteorologist has misread her notes and that the low was 1 degree above zero; or that the thermometer had given a false reading.

In each case, many background assumptions are in effect. At this point one might be tempted to opt for probabilities. But while the latter clearly have an important role to play in AI, they need not come in quite here. Instead, we often simply reserve judgment, or suspend a previous judgment.

And again, I-logics are vehicles for this real-time ongoing sort of reasoning. Indeed, an agent can only reason with what it has at hand.[10]

I-logic (at least in its active-logic form) not only brings many benefits but (perhaps surprisingly) is not particularly mired in the weeds of implementational details. This is not to say that all such issues are now fully resolved – this is a long work very much still in progress. But looking under the robotic hood, so to speak, is essential if we are to come to grips with how CSR can actually take place in robotic creations coming in the (seemingly quite near) future.

Thus instead of axioms, at any moment, our artificial agent has a specific collection of beliefs (stored in memory) and this collection changes as inferences are drawn, perceptions made, and so on. Among these changes – and central to most of the distinct features of active logic – is the updating of the present time as in the clock rule. There is no notion of inferential closure; the current beliefs are simply whatever has been inferred/perceived and kept so far (i.e., inherited to the present time).

A belief can fail to inherit for a variety of reasons. No belief of the form Now(t) is inherited – it is replaced by Now(t+1). Other failures of inheritance are illustrated in various cases below. But more importantly we now turn to the power of introspective reasoning that becomes possible in I-logics endowed with a notion evolving time.

Introspection Is a Many-Splendored Thing

Introspection is one of the most valuable tools that come almost for free in an I-logic.[11] It in turn facilitates powerful methods for detecting and defusing contradictions, managing nonmonotonic inference, reasoning about and adjusting semantics, tracking actions, and much more. In this and several sections that follow, we explain and illustrate a number of these ideas.

Given a belief P at time t, an agent ought to be able to note later on (say at time t+1) that it had that belief earlier. This can be achieved in active logic by means of a rule such as the following (positive introspection), where the KB-

---

[9] This is sometimes embraced as a necessary evil (Reiter 2001); or dealt with via specialized semantics (Levesque&Lakemeyer 2000) which however does not adequately address or ameliorate the time-consumption aspect.

[10] See for instance the Oxford Reference on *Neurath's boat* – "The powerful image conjured up by Neurath, in his Anti-Spengler (1921), whereby the body of knowledge is compared to a boat that must be repaired at sea: 'we are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom…'. Any part can be replaced, provided there is enough of the rest on which to stand. The image opposes that according to which knowledge must rest upon foundations, thought of as themselves immune from criticism, and transmitting their immunity to other propositions by a kind of laying-on of hands."

[11] And so perhaps "introspective logic" would be a more apt name than internal logic.

predicate symbol refers to the agent's own knowledge base:

```
t:    P
---------------
t+1: KB(P,t)
```

Similarly, another rule (negative introspection) can provide the result that one did not just previously have a given belief:[12]

```
t:    …
-----------------
t+1: ~KB(P,t) [if P is not present at the previous step]
```

These two rules are trivial to implement and cheap to run, involving no more than a linear-time lookup at time t+1 to see what wffs are or are not among the t-beliefs.[13] Yet a surprising number of capabilities flow from this, as expanded upon in the next several subsections.

Non-monotonicity
At this point we can already carry out some simple cases of nonmonotonic reasoning. For instance, the default that B's are typically F's (birds typically fly) can be captured like this: if one doesn't already (as in a moment ago) know that a given bird doesn't fly, then assume it does. In active-logic notation this can be written as follows:

$$(\forall x) \ [\ (\forall t) \ \{Bird(x) \ \& \ \sim KB(\sim Flies(x),t\text{-}1)\} \rightarrow Flies(x) \ ]$$

Then given Bird(tweety), all it takes to infer that Tweety can fly is ~KB(~Flies(tweety),t-1), which comes instantly from negative introspection – unless one does already know Tweety cannot fly. No fuss, no muss – no need for complex consistency checks or internal model-building; conclusions are held as long as they are held, and can be surrendered when evidence so suggests.[14]

Thus, one might later on come to believe Tweety is a penguin – whether by observation or simply additional inference. This will then appear as a (direct) contradiction in the KB: two beliefs of the form P and ~P will both be present at the same time-step. Which brings us to the next subsection.

Contradictions
Contradictions are virtually inevitable in commonsense reasoning (Perlis 1997). While this is generally considered a major nuisance for CSR, it can actually be a boon. Here

is how an I-logic can benefit (in the specific form of active logic): If the wffs P and ~P both appear as t-beliefs, then neither are inherited as (t+1)-beliefs and instead Contra(t, P, ~P) is inferred as a (t+1)-belief. Thus the agent retains in the evolving present the fact that there had been an earlier contradiction, but is no longer directly subject to it, and *ex contradiction quodlibet* (from a contradiction all follows) is thereby disarmed.[15]

Thus instead of being a logician's anathema, contradictions can be a robot's best friend, helping it adjust its KB to come more into line with reality. Contradictions simply remain undiscovered in the KB until they *are* discovered (in the P, ~P form) over time – and then defused. This is a very different approach from more customary paraconsistent logics, most of which skirt around the edges of a contradiction – rather than acknowledge it and use it to make changes to the KB – or in effect assume they can all be hunted out in advance.[16]

In the case of Tweety above, new information that she is a penguin and does not fly will provide (say at time-step t) a direct contradiction between Flies(tweety) and ~Flies(tweety), which then at time t+1 will result in the KB having neither of these inherited from step t, but instead will have an assertion that such a contradiction did arise at time t. If the agent has further information – such as that penguins are a subclass of birds, and that subclass properties are more trustworthy[17] – then ~Flies(tweety) can be reinstated. If not, then the agent remains in doubt.

It is our contention that this sort of fluctuating conflict-resolution over time is the only option for an actual agent engaged in reasoning as the world evolves.

Semantics and Pragmatics
In an I-logic, semantics can take on an entirely new aspect, where the agent can exert control and both determine and reason about what its expressions do or don't stand for.[18] This is one of the most powerful aspects of introspection that we have noted so far. In effect, one can reason about one's own expressions – simply by means of introspectively examining past beliefs and subexpressions thereof. One

---

[12] Many issues arise here that we do not have space to address, such as: to which wffs P are the introspection rules applied (if care is not taken, the KB will quickly become swamped). A much longer paper in preparation will deal with this.
[13] A t-belief is simply any belief in the KB at time t.
[14] Of course, an agent can also remain in doubt, or even be deliberately tentative (such as with probabilities and during learning; see (Getoor&Taskar 2007)).

[15] To be sure, whatever circumstances that produced P and ~P may do so again, so this is not a panacea. But it can be shown (Miller 1993) that under reasonably broad conditions this too will resolve into a stable state.
[16] E.g., see (Roos 1992) for a more traditional E-logic treatment; and (Anderson et al 2013) for more on an active logic approach.
[17] Such a rule has been implemented in one of our active logic programs.
[18] That is, this refers to meanings the *agent* assigns to its expressions, quite apart from what a logic-designer may have in mind. Note that the recent Facebook robot-incident of "inventing a new language" is not of this sort at all: those robots did not assign meanings to anything, either in the original English or in their later made-up phrases. See (http://www.newsweek.com/2017/08/18/ai-facebook-artificial-intelligence-machine-learning-robots-robotics-646944.html)

can even assign new expressions, if for instance a new entity is observed, or if one infers that two entities were being conflated as one (as in the cases of ambiguity or of misidentification).

In fact, AI systems are generally notorious for altogether ignoring the expression/meaning distinction, as in: Joe is a person and also we just now used "Joe" to refer to him. People can and do (and must) note and make use of the difference between language and what language refers to. Our artificial agents need to be able to do the same; otherwise they can hardly be said to know anything (Perlis 2016), let alone reason about errors. With all the recent successes in NLP (mostly coming from deep learning), still there is almost no language-like introspection, no meanings associated with words in a way that allows reasoning, let alone adjusting meanings.

On the other hand, introspection allows representation of beliefs (at least at previous steps) as objects that can be reasoned about. This has numerous ramifications, which for lack of space we can only briefly allude to in the rest of this section.

Ambiguity and Misidentification

A potentially ambiguous expression (say, "Jean's car") can be recognized as such (e.g., by noticing a direct contradiction – "this is Jean's car, and the key to Jean's car isn't the key to this car"). This in turn triggers an effort to resolve the contradiction. Maybe Jean has two cars (ambiguity); or maybe this is the wrong key or that is not her car at all (misidentification).

The latter case is especially interesting, for it requires *some* expression to represent an object (the wrong key or wrong car), but not the expression that had been used a moment ago. Miller and Perlis (Miller&Perlis 1996) propose a special active-logic function-symbol *tfitb* to produce a new name on demand, for the "thing formerly interpreted to be" something else.

Focal points

A related idea comes up in planning, especially multiagent planning. It may be important to identify an entity that another agent is likely to similarly identify – for instance a good location to meet up or to leave a message, or an "obvious" item to pick out of a long list (e.g., the first, last, or middle one). This in turn may require coming up with a new expression that was not previously in one's ontology. In (Kraus et al 2000) an approach to this is given using active logic.

Pragmatics

In conversation, all sorts of assumptions arise and are confirmed or dispelled, often by means of further conversation. Thus NLP-dialog is a prime example of beliefs com-

ing and going during reasoning. Here is one example dialog, in which reasoning involves inferences that evolve over time, that has been implemented in active logic (Purang et al 1996):
(A) Kathy: Are the roses fresh?
(B) Bill: They are in the fridge.
(C) Bill: But they are not fresh.
At some point prior to (C), Bill supposes Kathy will draw from (B) the implicature that the roses are fresh, so in (C) he dispels that inaccuracy. Thus Bill has to reason about the effects of the ongoing conversation and make adjustments to it.

The One Wise Man Problem

Much has been made of the Three-Wise-Men problem – see (Konolige 1984; Elgot-Drapkin&Perlis 1990). A realistic treatment has to take into account the passage of time as the wise men think; and this can be done in traditional temporal logic, as long as the wise men themselves are not required to use that same logic. But suppose we do want to capture the reasoning of such an agent; for instance – to make the problem especially simple – the King who wants to assure himself that his one wise man is not an idiot. So the King proposes this problem to his wise man: "Is 15 a prime number?" Being no genius himself, the King has to think for awhile before deciding the answer is "no" – and if by then the wise man has not yet answered, the King can start looking for a replacement. But to do this reasoning (which involves introspection), the King will need I-logic, and in particular an I-logic that closely tracks time.

What Am I Doing?

It is important that an agent not only plan and take actions, but that it also know when it is in fact doing so. Otherwise strange behaviors can result. In one of our robotic studies recently, robot Alice was programmed to point and say "I see Julia" whenever it heard an utterance containing the word "Julia" (actually, it was doing no actual word-processing at the time, but simply matching the input sound-stream to a stored one). So it got itself into a loop, hearing "Julia" from its own loudspeaker and then pointing and repeating the same phrase over and over.

But taking a cue from neuropsychology,[19] we were able to encode a rule for noting one's own activity: whenever an action is undertaken, Do(x) is inferred (recall the Lunch example), and at the next step Doing(x) can be inferred, and inherited as long as the activity is still underway.[20] We have implemented this in a grounded way, so that when Alice undertakes to speak she infers that she is engaged in

---

[19] The so-called *efference copy*, see (Brody et al 2015).
[20] This is a different method from that used in (Bringsjord et al 2015) where voice recognition appears to take precedence over recall of one's own actions.

a speaking action (but also checks what she hears to make sure it matches her expected speech).

Reasoned Learning

Machine learning (ML) has taken center stage in recent years, and for good reason: it has made justly fabled strides, and surely will be a major part of any future general-purpose AI. But alone it is insufficient. The practices usually referred to as ML are ones of habituation or training. A human turns a trainable system on, allows it to train, perhaps applies it, and later turns it off; in itself, traditional ML has little if any autonomy.

But a general-purpose AI (robotic or otherwise) will need to decide what to learn, and when and how, and whether learning is working and/or should stop. Moreover, as noted in the Introduction, cultural (symbolic) transmission is also a major source of learning.[21] And finally, a system will need to know what it has or hasn't already learned.[22]

An I-logic (particularly, active logic) – in keeping a history of its own KB over time – can potentially examine that history, infer that it has (or lacks) certain capabilities, and then decide whether to activate an appropriate ML process; see (Elgot-Drapkin, et al 1991) for a brief introduction.

Related Work

Ray Reiter (Reiter 2001) considers numerous issues that arise in commonsense reasoning (CSR) when an agent's deliberations occur within a dynamic setting, and in particular, how a formal logic might be used *by* an agent to do its own reasoning, and have that reasoning keep up with changing events (pp.163-164). Reiter succeeds in isolating various themes surrounding this: omniscience, internal contradictions, and so on. But in the end he advocates instead the "external design stance." *Action languages* (Gelfond&Lifschitz 1998) are another firmly E-logical approach that thus again are suitable for external analysis of an agent but not for real-time use by an agent, let alone by one with a potentially inconsistent KB; the same holds for *temporal action logics* (TAL; see Doherty 1998) and *the temporal logic of actions* (TLA; see Lamport 1994).

In a survey of commonsense reasoning (Davis 2017) the E- and I- distinction is also raised (under different terminology); but, like Reiter, he focuses primarily on the external stance. A survey on robot deliberation (Ingrand&Ghallab 2017) does not address this distinction.

Levesque and Lakemeyer (Levesque&Lakemeyer 2000, pp 195-196) argue that attending to internal inference mechanisms to avoid omniscience makes behavioral predictions impossible. They deal with omniscience instead by enlargements of the semantics to allow "non-standard world states" that keep out undesired agent-beliefs. But it is unclear what predictions one could hope to make, given an agent with thousands of explicit beliefs, other than ones of such generality as to be virtually useless about that particular agent's behavior. Will it complete a given task (even a purely inferential one) within ten days? One surely cannot expect anything other than a careful examination of the robot's actual processing to reveal such results.

On the other hand, Richard Weyhrauch and Carolyn Talcott (Weyhrauch 1980; Weyhrauch&Talcott 1990, 1994; Talcott 2003) initiated the FOL approach (one instance of an I-logic) which aimed at providing reasoning mechanisms for actual use by an agent; however this effort has remained in a fragmentary state. An interesting addendum to FOL is WristWatch (Weyhrauch & Talcott 1997)—a dynamic context from which to answer questions about time, specifically about the ever-changing meanings of the constants *now* and *then* as updated by their "tick" inference rule. Weyhrauch and Talcott speculate about supplying a robot with WristWatch embedded into FOL as its mechanism to reason about time.

Pei Wang's Non-Axiomatic Logic (aka NARS) provides a (term-logic based) reasoning system which aims to be finite, real-time and open (Wang 2013). It shares some features with active logic, in that it is non-monotonic, allows for self-reference and is intended to be situated (in that knowledge is not disembodied but should be based on the agent's experience). While Chapter 9 of (Wang 2013) addresses potential meta-cognition in his system, no particular mechanisms for monitoring an ongoing reasoning process seem to be specified. Gestures toward such mechanisms are made (by, e.g., referencing "doubt" and "wait" operations), but we are not aware of any attempt to operationalize these. Later iterations of NARS (Wang&Hammer 2015; Hammer et al 2016) address temporality and recognize the problem of assuming that "the reasoning system itself is outside the flow of time" (Wang&Hammer 2015). The temporality in this system differs from active logic, however, in that the flow of time is not itself seen as an object of reasoning.

Jacek Malec and his group (Asker&Malec 2005) extended active logic and proposed a labeled deductive system (LDS) which attaches a label to every well-formed formula. LDS allows the inference rules to analyze and modify labels, or even trigger on specific conditions defined on the

---

[21] See also (Levesque 2017).
[22] But again see (Getoor&Taskar 2007) for another approach.

labels. They demonstrated the use of LDS by formalizing models of short-term memory, followed up by studying several scenarios (Heins 2009). In related work, Nowaczyk 2006) extends active logic to partial planning situations.

An interesting middle-ground is taken in TRL – *timed reasoning logic* – see (Alechina et al 2004a,b; Agnotes&Alechina 2007). While TRL remains at the E-logic level, it can express fairly detailed aspects of internal processing. In that respect it is similar to the meta-level step-logics in (Elgot-Drapkin&Perlis 1990). Because of more limited expressive power, TRL tends to be decidable. On the other hand, the semantics given in (Anderson, et al 2008) appears to offer a compelling psychologically plausible alternative. But it is noteworthy that none of these address the agent-controlled-semantics issue above.

The planning community is beginning to acknowledge the importance of taking planning-time into account as part of the planning process; see for instance (Ghallab et al 2016; Lin et al 2015). The earliest published work we are aware of on this is (Nirkhe et al 1991).

A recent article (Tenorth&Beetz 2017) discusses complex interactions between robotic control, knowledge representations at various levels, and reasoning over those representations, including temporal reasoning. While the intention is to provide robots with inferential abilities, the approach appears to remain in the E-logic framework.

Conclusion: Reasoning is a Process

A reasoner is *engaged* in reasoning, and makes decisions *during* (and as part of) that reasoning, such as whether to continue along present lines, or try a new tack, or give up, or seek assistance. That is, a reasoning agent itself is engaged in some version of what we have called I-logic. On the other hand, the *study* of reasoning can of course proceed at many levels and in many forms.

It may be premature – despite many decades of work (including some by ourselves) – to try to pin down precise specifications (i.e., in an E-logic) of broad CSR behaviors. We know so little of the notion of intelligence at this point, that it may be more useful to get lots more experience with reasoning behavior itself (that is, via I-logics that can actually be used by automated agents/robots). At least, this is the perspective we are exploring here.

An analogy with (Polya 1945) is tempting. While mathematical logic is the very epitome of E-logic (fully focused on entailment/consequence), it largely ignores the situation of actual mathematician-reasoners who question axioms,

decide to change problems, and are keenly aware of (and make use of) their progress or lack of it over time (Perlis, 2016). Polya's advice is aimed at the latter, with practical in-the-moment strategies to attend to. And while mathematical logic has been extraordinarily successful in its own right, it has afforded relatively mild impact or insight into mathematical practice overall.

We repeat from our Introduction: The single most salient departure that I-logics make from E-logics is that of *taking into account the actual process of inferring as something that itself takes time*. This departure provides a very rich set of tools that we hope to have illustrated here.

References

Agnotes, T. and Alechina, N. 2007. The dynamics of syntactic knowledge. *Journal of Logic and Computation*. February 2007

Alechina, N., Logan, B., and Whitsey, M. 2004a. A complete and decidable logic for resource-bounded agents. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems* (AAMAS 2004). ACM Press.

Alechina, N., Logan, B., and Whitsey, M. 2004b. Modelling communicating agents in timed reasoning logics. *Proceedings, 9th European Conference* (JELIA) – Lecture Notes in AI 3229, Springer.

Anderson, M., Gomaa, W., Grant, J., and Perlis, D. 2013. An approach to human-level commonsense reasoning. In K. Tanaka, F. Berto, E. Mares, and F. Paoli (eds.), *Paraconsistency: Logic and Applications*. Springer.

Anderson, M., Gomaa, W., Grant, J., and Perlis, D. 2008. Active logic semantics for a single agent in a static world. *Artificial Intelligence* 172: 1045-63.

Anderson, M. and Perlis, D. 2005. Logic, self-awareness and self-improvement: The metacognitive loop and the problem of brittleness. *Journal of Logic and Computation*. . 15(1)

Asker, M. and Malec. J. 2005. Reasoning with limited resources: Active logics expressed as labeled deductive systems. *Bulletin of the Polish Academy of Sciences, Technical Sciences*, Vol. 53, No. 1, 2005, pp. 69-78.

Baral, C. and Jicheng Z. 2008. Non-monotonic temporal logics that facilitate elaboration tolerant revision of goals**.** AAAI 2008: 406-4

Barringer H, Fisher M, Gabbay DM, Gough G, editors. 2013. *Advances in temporal logic*. Springer Science & Business Media; 2013 Nov 11.

Bringsjord, S., Licato, J., Govindarajulu, N., Ghosh, R., and Sen, A. 2015. Real robots that pass human tests of self-consciousness. In Proceedings of IEEE the 24th International Symposium on Robots and Human Interactive Communications

Brody, J., Cox, M.T., and Perlis, D. 2014. Incorporating elements of a processual self into active logic. In M. Waser (Ed.), *Implementing selves with safe motivational systems and self-improvement*: Papers from the Spring Symposium, AAAI Press.

Brody, J., and Perlis, D. 2015. Who's talking? Efference copy and a robot's sense of agency. AAAI Fall Symposium 2015, Arlington VA, 2015

Davis, E. 2017. Logical formalizations of commonsense reasoning: a survey. *J. Artificial Intelligence Research*, 59, 651-723.

Dell'Aglio, D., Della Valle, E., Van Harmelen, F., and Bernstein, A. 2017. Stream reasoning: A survey and outlook. *Data Science,* in press.

Delgrande, J., Peppas, P., and Woltran, S. 2013. AGM-Style Belief Revision of Logic Programs under Answer Set Semantics. LPNMR 2013: 264-276

Diller, H., Adrian Haret, Thomas Linsbichler, Stefan Rümmele, Stefan Woltran, 2015. An Extension-Based Approach to Belief Revision in Abstract Argumentation**.** IJCAI 2015: 2926-2932

Doherty, P. 1998. TAL| Projects| AIICS| IDA. *Electronic Transactions on Artificial Intelligence*, *2*(3-4), 273-306.

Elgot-Drapkin, J., and Perlis, D. 1990. Reasoning situated in time, I: basic concepts. *J. of Experimental and Theoretical AI.*

Elgot-Drapkin, J., Miller, M., and Perlis, D. 1991. Memory, reason and time: the step-logic approach. In R. Cummins and J. Pollock (eds): *Philosophy and AI: Essays at the Interface.* MIT Press.

Gärdenfors P, 2003. (editor) *Belief revision*. Cambridge University Press.

Gelfond, M. and Lifschitz, V. 1998. Action languages. *Electronic transactions on AI*, v.3.

Getoor, L, and Taskar, B. (editors) 2007. *Introduction to Relational Statistical Learning*. MIT Press.

Ghallab, M, Nau, D., and Traverso, P. 2016. *Automated Planning and Acting*. Cambridge Univ. Press.

Goldsmith, J., Sloan, R.,Szorenyi, B., and Turán, G. 2004. Theory revision with queries: Horn, read-once, and parity formulas**.** *AIJ 156(2):* 139-176

Gonzalez, G., Baral, C., and Cooper, P. 2002. Modeling multimedia displays using action based temporal logic. VDB 2002: 141-155

Hammer, P., Lofthouse, T., and Wang, P. 2016. "The OpenNARS implementation of the non-axiomatic reasoning system." International Conference on Artificial General Intelligence. Springer International Publishing

Heins, T. 2009.. A case study of active logic. Master's thesis, Department of Computer Science, Lund University.

Ingrand, F. and Ghallab, M. 2017. Deliberation for autonomous robots: A survey. *Artificial Intelligence,* v 247.

Konolige, K. 1984. *A deduction model of belief and its logics.* PhD dissertation, Stanford University.

Kraus, S. and Lehmann, D. 1986. Knowledge, Belief and Time. ICALP 1986: 186-195

Kraus S., Rosenschein, J. S., and Fenster, M. 2000. Exploiting focal points among alternative solutions: two approaches. *Annals of Mathematics and Artificial Intelligence*, 28(1-4):187-258.

Lamport, L. 1994. The temporal logic of actions. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, *16*(3), 872-923.

Levesque, H. 2017. *Common Sense, the Turing Test, and the Quest for Real AI*. MIT Press.

Levesque, H., and Lakemeyer, G. 2000. *The Logic of Knowledge Bases*. MIT Press.

Lin, C., Andrey, K., Ece, K., and Horvitz, E. 2015. Metareasoning for planning under uncertainty. arXiv:1505.00399v1

Miller, M. 1993. *A view of one's past, and other aspects of reasoned change in belief.* PhD dissertation, University of Maryland.

Miller, M., and Perlis, D. 1996. Automated inference in active logics. *J Applied Non-classical Logics, 6(1).*

Mueller, E. 2015. *Commonsense Reasoning*, 2nd edition. Morgan Kauffman.

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

Nirkhe, M., Kraus, S., and Perlis, D. 1991. Fully deadline-coupled planning: One step at a time. International Symposium on Methodologies for Intelligent Systems (ISMIS 1991).

Nowaczyk, S. 2006. Partial planning for situated agents based on active logic. *Workshop on Logics for Resource Bounded Agents, ESSLLI*

Perlis, D. 1997. Sources of, and exploiting, inconsistency: preliminary report. *Journal of Applied Non-Classical Logics*.

Perlis, D. 2000. The role(s) of belief in AI. In J. Minker (ed.) *Logic-Based AI*, Kluwer.

Perlis, D. 2016. The five dimensions of reasoning in the wild. AAAI-2016.

Pnueli A. 1977. The temporal logic of programs. In *Foundations of Computer Science,* 1977., 18th Annual Symposium. Oct 31 (pp. 46-57). IEEE.

Polya, G. 1945. *How to Solve It*. Princeton Univ. Press.

Purang, K., Gurney, J., and Perlis, D. 1996. AAAI Spring Symposium.

Rajan, K. and Saffiotti, A. (editors) 2017. Special issue on AI and robotics. *Artificial Intelligence, v 247*.

Reiter, R. 2001. *Knowledge in Action.* MIT Press.

Roos, N. 1992. A logic for reasoning with inconsistent knowledge. *Artificial Intelligence*, 57, 69-103.

Sloan, R. and Turán, G. 1999. On theory revision with queries**.** *COLT* 1999: 41-52

Talcott, C. 2003. FOL: Towards an architecture for building autonomous agents from building blocks of first order logic. http://www-formal.stanford.edu/FOL/03jan-umd.ppt (Slides from talk at U Maryland.)

Tenorth, M. and Beetz, M. 2017. Representations for robot knowledge in the KnowRob framework. *Artificial Intelligence, v 247.*

Wang. P. 2013. *Non-axiomatic logic: A model of intelligent reasoning.* World Scientific.

Wang, P., and Hammer. 2015. "Issues in temporal and causal inference." International Conference on Artificial General Intelligence. Springer.

Weyhrauch, R 1980. Prolegomena to a theory of mechanized formal reasoning. *Artificial Intelligence,* 13, 133-170

Weyhrauch, R., and Talcott, C. 1990. Towards a theory of mechanizable theories: I FOL contexts – the extensional view. European Conference on Artificial Intelligence (ECAI).

Weyhrauch, R., and Talcott, C. 1994. The logic of FOL Systems: formulated in set theory. In *Logic, Language, and Computation.* 119-132.

Weyhrauch, R, and Talcott, C. 1997. WristWatch – an FOL theory of time. http://www-formal.stanford.edu/FOL/w.ps