# Time-Situated Agency: Active Logic and Intention Formation[1]

Michael L. Anderson[1], Darsana P. Josyula[2], Yoshi A. Okamoto[3], Don Perlis[1,2] (1)Institute for Advanced Computer Studies, (2)Department of Computer Science, University of Maryland, College Park, MD 20742 (3)Microsoft Corporation, One Microsoft Way, Redmond, WA 98052 {mikeoda,darsana,yoshi,perlis}@cs.umd.edu

**Abstract**

In recent years, embodied cognitive agents have become a central research focus in Cognitive Science. We suggest that there are at least three aspects of embodiment—physical, social and temporal—which must be treated simultaneously to make possible a realistic implementation of agency. In this paper we detail the ways in which attention to the temporal embodiment of a cognitive agent (perhaps the most neglected aspect of embodiment) can enhance the ability of an agent to act in the world, both in itself, and also by supporting more robust integrations with the physical and social worlds.

## 1 Three Aspects of Embodiment in Cognitive Science

Any implemented, interactive system is embedded in the world in at least three ways: temporally, physically and socially.[2] Its processes take time and unfold within time; likewise, the system takes up space by (at least) being instantiated in and utilizing components with discreet and limited storage locations, and perhaps by being tightly and exclusively bound to a particular physical object (a robot being the most salient example); finally, insofar as it can at least receive commands and/or display results, it exists within and is oriented toward an interpretive, social world, a world not of objects but of other agents. It is well known by now that work in Cognitive Science in general, and in the new robotics movement in particular, has been rethinking the significance of the physical embodiment of cognitive systems: the physical space in which a system exists, and the physical object in which it is instantiated, are not merely limitations to work within, but represent a resource to exploit. Thus a great deal of attention has been paid to systems which are aware of, prepared to utilize, to treat as relevant, take advantage of, be affected by, and even change their physical environment in support of its goals and processes.

We believe that the same kind of re-thinking must take place for the temporal and social embeddedness of cognitive systems.[3] That reasoning takes place in time is not an obstacle the impact of which is to be minimized, but is a fact which can be recognized and utilized by a reasoning agent to improve its ability to think and act. Likewise, the necessity of interacting with the social world should not be thought of as an undesirable difficulty, but

[1]This research was supported in part by the AFOSR and ONR

[2]These are not the only important aspects of embodiment, nor even the only ones relevant to agency. Emotion and mood come immediately to mind as further aspects of embodied agents which must be considered for any fully realistic model of human agency. See [45, 13]

[3]Indeed, we believe attention paid to these issues can be used to improve interactive systems more generally. See, e.g. [9, 10, 11, 31, 32, 40, 49]

as an opportunity to access and utilize resources which can dramatically improve a system's power and performance.

In general, as Ismail and Shapiro succinctly state, the goal is to "use reasoning in the service of acting, and acting in the service of reasoning" [30]. We believe that this insight, which has driven so much interesting work in the physical embodiment of computational systems, can be fruitfully applied as well to both their temporal and social embodiment.[4] Through our own focus on temporal integration, we are trying to move forward on all three fronts simultaneously. We believe, for a system which reasons in time, that the temporal aspects of its reasoning are salient (a belief may have been held in the past, but no longer; a train of thought may be pursued for a long time without results; a decision may need to be made in light of an approaching deadline) and that therefore any such system should be made aware of the passage of time as its reasoning proceeds. Our work with active logics (section 4) demonstrates that greater temporal integration along these lines can improve reasoning systems, and in particular that these improvements can enhance the ability of a reasoning system to integrate with changing real-world environments, as well as to better understand and integrate with human users through natural language. This is the general case: the current essay highlights the importance of these types of integration to the formulation and implementation of intentions in the case of an autonomous agent acting in the world, and we direct the reader to other published work detailing other applications of temporally situated reasoning, as in the more restricted case of a servitor which must understand and implement a user's command.

# 2 Some Aspects of Agency

These three aspects of embodiment are especially salient for agency.[5] Roughly speaking, to intend an effect it is necessary to be able to identify an object, know its current state, and imagine (or otherwise represent) an alternative. To be able to intend *effectively* it must also be possible to map a path from the current to the intended state, and be in a position to implement the required changes; more than this, the integration between the planning/implementation system and the intentional system must be such that information about the possibility of an intention can be taken into account in the intention-forming process, and information about an agent's progress toward a goal can be taken into account in further (or ongoing) planning aimed at reaching the goal (effecting an intention).[6]

From its root, then, intending is bound up with temporal considerations, for it is of necessity future-directed, depending on a recognized difference between the known present, and a desired future; further, planning has a temporal structure, for some things must be done

[4] In [30] Haythem Ismail and Stuart Shapiro suggest a similar (but less broadly stated) philosophy to guide work on embodied cognitive agents.

[5] The discussion of agency is influenced throughout by [6, 16, 8].

[6] There are many senses of the word "intention"; we are most interested in that used in connection with those explicitly represented and consciously available goals which play a central role in acting (and planning to act) in the world. It is worth distinguishing this in particular from a sense of "intended" more closely bound to the notion of voluntary. See [6] for an excellent discussion.

first, and others last, some now and others later, and there may be groups of things that, while they must be done sequentially with respect to each other, are as a group temporally and conditionally independent of the other parts of a complex action. This feature of agency is sometimes called a plan hierarchy, and it might be thought that the relations can be characterized in a time-independent way, in terms of conditional dependencies and prerequisites. Even where such re-description is possible, however, it seems that it would submerge real features of human agency and planning, for in designing and implementing a complex plan, it will be important to know what sub-plans will take longer, (and so must be started now, despite conditional independence from other parts of a plan), which can be done simultaneously, and which would be best to *finish* simultaneously. That is, planning seems to involve considerations of *timing* which cannot be expressed purely in terms of functional dependence. Obviously, agency also requires awareness of environment, and of the agent's own situation in that environment; and in many cases agency is also importantly bound up with social awareness, for an intention can originate in a request or command, or it could require the cooperation of other agents. In a more complicated case an intention could have ethical dimensions which would play a role in determining the methods whereby (or even whether) it was effected.

The intentional system also has its own internal structure or hierarchy which is important to its nature. There are at least two axes to this structure, one temporal and one directional. The temporal axis is structured by the difference between short term (proximal) and long term (distal) intentions. We might get at the difference between proximal and distal intentions by saying: if intention Q could be provided by an agent in justification or explanation of the action P, and if action P, once completed, does not fulfill the intention Q (if it remains rational to intend Q), then Q is a distal intention. Thus, the intention Q (so Charles will break his leg) provided in explanation of action P (putting a roller skate in the middle of the hall) is distal. In contrast, going upstairs to get the camera represents a proximal intention. An action can be distally related to an intention for two different reasons: first, it can be reasonably explained by Q only in light of a *prediction* R (that Charles will come through the door and step on the skate, etc.); second, it can be one action in a planned series of actions, such that the series S, but not the action P, will result in the fulfillment of Q. Another way to get at the distinction is in terms of whether a given intention seems to require distinct other intentions for its fulfillment. Getting the camera, despite the fact that doing this requires going up the stairs, into the bedroom, grabbing the camera (and that these actions would be intentional in the sense of voluntary, mentioned above), doesn't seem to require further explicit intentions or predictions.[7] On the other side of the coin, "writing my dissertation", offered in explanation of typing away on my computer, looks to express a distal intention; there is no single, coherent action, unguided by further explicit intentions, which constitutes "writing one's dissertation".[8]

[7] This assumes an obvious distinction between predictions, which are not required, and expectations (that the camera is upstairs), which are.

[8] There is a complication here, because actions themselves can be more and less narrowly defined; there will almost always be a way to construe P so as to make it distally related to any given Q. This is true, but perhaps not telling, for there will generally be a way to construe Q so as to make it proximal with respect to P. This will be the case whenever there is a single coherent (not necessarily simple) action which constitutes "the doing of Q".

It is not important that every case be clear; the image of a structural axis suggests a continuum, where some intentions are clearly proximal, and others clearly distal. However, that there is such a structure to the intentional system *is* important, and it is one of the insights which lies behind Michael Bratman's notion of a *planning agent*[8], which we might describe as an agent which is capable of analyzing distal intentions into proximal ones, and proximal intentions into actions.

Bratman describes planning agency this way:

> Our purposive activity is typically embedded in multiple, interwoven quilts of partial, future-directed plans of action. We settle in advance on such plans of action, fill them in, adjust them, and follow through with them as time goes by. We thereby support complex forms of organization in our own, temporally extended lives and in our interactions with others; and we do this in ways that are sensitive to the limits on our cognitive resources.([8] p.1)

In addition to the welcome focus on the temporal aspects of agency, the central upshot of this picture of agency is that committing one's self to a (distal) intention is more like driving a car to a destination than throwing a ball at a target—one must continually observe one's progress to the adopted end, and use these observations to make decisions and adjustments to best guide one's actions.[9]

Further, it is important to monitor the world and our effects on it not just in the service of guiding current intentions to their fulfillment, but also in the service of maintaining an accurate self-conception. It seems clear that an important part of our ability to intend, to choose means, and to guide actions to their ends is bound up with an accurate assessment of our particular abilities and capacities as practical agents. Just as is the case with knowledge of the world, this self-knowledge must be constantly monitored for accuracy. Thus, our ability to maintain an effective practical agency requires not just reasoning about actions in light of the world, but reasoning about ourselves (and our beliefs about ourselves) in light of the success or failure of our actions; in other words, agency requires not just reasoning, but introspection and meta-reasoning as well.

This brings us to the directional axis of the intentional system, which concerns the object at which the intention aims: there are internally directed intentions, which only effect other intentions, and externally directed ones, aimed at the world. An example of the first sort might be an intention about what sort of agent (person) to be [22]; such an intention needn't be anything so grand as wanting to be a moral saint, but can involve more mundane matters like a desire for efficiency. A desire to be an efficient person will figure in decisions about which of many possible paths to follow to a given end; in a more complex case it could figure even in choosing which strategy of *reasoning* to use to decide between competing paths (see

[9] It is worth emphasizing what is only mentioned in the quote, but is centrally important to Bratman's theory of agency—that we are not just individually planning agents, but cooperative ones. We not only plan complex projects in coordination with one another, but even in the pursuit of individual intentions consider ways in which we might secure each other's cooperation.

[14] for a brief discussion). This points to another kind of metareasoning required of a fully-specified autonomous agent.

# 3 Agency and Uncertainty

Intention formation—indeed, planned, directed action more generally—requires continual observation, cooperation and planning; in addition a robustly specified agent seems to require introspection and meta-reasoning. As has been mentioned already, a good deal of the current work most concerned with robust physical/environmental integration adopts a highly *reactive* model of agency, designed to produce complex behaviors without detailed internal representations; the field in general is re-thinking the more traditional "symbolic processing" approach to modeling and producing intelligent behavior.[4, 2] However, in so far as the above analysis is correct, agency requires not just continual observation of and reactive adjustments to the physical environment, but also introspective observation and the ability to engage in meta-reasoning about intentions and capacities, making appropriate *internal* adjustments. This suggests that a more complete agent must be both reactive *and* deliberative; yet these are generally considered antithetical goals. One (but certainly not the only) way in which this tension can be brought out is by considering the problem of reasoning under uncertainty.

The real world is complex, dynamic, and not completely knowable. What is true now may not have been true before, or for much longer, and more can always be discovered. Any system that purports to model the world (or any part of it) must be able to accommodate such changes. Given these conditions, any reasoning about the world is provisional and uncertain, because changes in what is true of, or known about the world can require revisions to simple beliefs, derived conclusions, generalizations, and even heuristics for thinking. Systems that hope to accomplish this latter task must be flexible enough to recognize and gracefully handle these situations, and recover from the contradictions, inconsistencies, and irregularities that they involve. Much the same can be said about integration with the social world. Focusing just on linguistic interaction (which must be considered central to social integration), complexity and uncertainty seem the salient characteristics. Conversation is not generally the exchange of fully formed, grammatically correct, and error-free utterances. Indeed, it is unlikely that there could ever be a fully fluent, error-free dialog; even putting aside problems of signal reception, and assuming perfect syntactical processing, the ability to understand a dialog partner involves such complicated and uncertain tasks as modeling their knowledge state and using context to disambiguate reference.[10] As in the case of observing the physical environment, one must be prepared to retract conclusions, and engage in repairs of one's beliefs as well as of the dialog itself, as the conversation continues and more evidence comes

[10] This latter task points already to ways in which social and physical integration are intertwined, and suggests the sorts of reasons one might give for moving forward with these two aspects of embodiment simultaneously: for disambiguating a reference ("I guess he's had enough.") can require not just attention to the *dialog* context (in which "he" might refer to the current subject of conversation), but also to the *physical* context of the dialog (in which "he" might be taken to refer to the fellow who just fell off the barstool).

to light.

Thus, it looks like a deliberative agent will require robust representations and principled methods for deriving further beliefs and adjusting current beliefs, while a reactive agent is guaranteed to confront incoherent, contradictory, or otherwise flawed information in the course of its interaction with the world. But, and here's where the tension between these two modes is most evident, from a contradiction everything logically follows.[11] More technically, from a direct contradiction, $P\&\neg P$, all well formed formulas are entailed as theorems. Clearly, a reasoning system which will come to believe everything it is possible to conceive will not be useful to a real-world agent.

What is wanted then, for implementing real-world agency, is a model of (logical) reasoning that can:

(1) A. continue to reason in the presence of contradictions (since they are inevitable)

B. be able to detect contradictions, curtail nonsensical inferences, and initiate repairs, and

C. support rich representations sufficient for robust meta-reasoning.

Active logics are designed to meet these desiderata and we believe they can be used to provide on-board reasoning capability for a real-time agent.

# 4 Active Logic: An Introduction

Active logics are a family of formalisms that combine inference rules with a constantly evolving measure of time (a 'now') that itself can be referenced in those rules. An account of the basic concepts can be found in [21].

One of the original motivations for active logics was that of designing formalisms for reasoning about an approaching deadline; for this use it is crucial that the reasoning takes into account the ongoing passage of time as that reasoning proceeds. Such a formalism has the ability to explicitly track the individual steps of a deduction, making it a natural mechanism for reasoning about contradictions and their causes.

Each "step" in an active logic proof itself takes one active logic time-step; thus inference always moves into the future at least one step and this fact can be recorded in the logic. The KB will at all times be finite since the finitely-many inference rules can produce only finitely-many conclusions in one time-step.[12] The meaning of an inference rule such as 2 (an active logic analogue to *modus ponens*), is that if $A$ and $A \rightarrow B$ are in KB at time (step number) `i`, then $B$ will be added to the KB at time `i+1`.

[11] This is not the place to survey the various approaches to reasoning under uncertainty, nor to rehearse our reasons for pursuing logic-based reasoning methods. For more information on these matters see [20, 21, 36, 7].

[12] In ongoing work begun in [38] we have been exploring ways to keep the KB size not merely finite but bounded, analogous to human short-term memory.

$$(2)\quad \begin{array}{ll} \texttt{i} & : \\ \texttt{i+1} & : \end{array} \frac{\texttt{A},\ \texttt{A} \rightarrow \texttt{B}}{\texttt{B}}$$

(In general there may be conditions that must be met before such a rule can fire—see below; but if a rule can fire, it will.) In addition to the new formula B, the KB at step `i+1` would contain all the formulas that are inherited from step `i`. By default, all beliefs from one step that are not directly contradicting are inherited to the next step. However some beliefs like the ones related to the current time are not inherited to the next step. (See below). The inheritance of formulas from one step to the next is controlled by inheritance rules. One simple version of such an "inheritance rule", which also illustrates the use of firing conditions, is shown in 3:

$$(3)\quad \begin{array}{ll} \texttt{i} & : \\ \texttt{i+1} & : \end{array} \frac{\texttt{A}}{\texttt{A}}\ \texttt{[condition:}\ \ \neg\texttt{A} \notin \texttt{KB at step i and A} \neq \texttt{Now(i)]}$$

To achieve much of their reasoning, active logics employ a notion of "now" that is constantly updated by the "clock rule" shown in 4:

$$(4)\quad \begin{array}{ll} \texttt{i} & : \\ \texttt{i+1} & : \end{array} \frac{\texttt{Now(i)}}{\texttt{Now(i+1)}}$$

An active logic keeps track of the passage of time using the current value of "now", so it is important that the value of "now" from a previous step is not inherited to the next step. The firing condition in the inheritance rule in 3 would prevent `Now(i)` lingering in KB after step `i` along with the newly inferred `Now(i+1)` [13]

Theorems can be marked with their time (step-number) of being proven, i.e., the current value of "now". This step-number is itself something that further inferences can depend on, such as inferring that a given deadline is now too close to meet by means of a particular plan under refinement if its enactment is estimated to take longer than the (ever shrinking) time remaining before the deadline.

Active logic formalisms are distinct from traditional temporal logics, in that the latter characterize truth about past, present, and future as if from a timeless (or unchanging) present; that is, the inferences do not formally correspond to an increase in the value of "now". This is appropriate as long as the temporal reasoning is by one agent about another agent far removed in time, or if the latter agent's activity is independent of the former. But when an agent is reasoning about its own ongoing activity, or about another agent whose activity is highly interdependent, traditional "time-frozen" reasoning is at a disadvantage, and "time-tracking" active logics can bring new power and flexibility to bear.

It is the time-sensitivity of active logic inference rules that provides the chief advantage over more traditional logics. Thus, an inference rule can refer the results of all inferences *up until*

[13] Inheritance and disinheritance are directly related to belief revision [23] and to the frame problem [34, 12]; see [38] for further discussion.

*now*—i.e. thru step $i$—as it computes the subsequent results (for step $i+1$). This allows an active logic to reason, for example, about its own (past) reasoning; and in particular about what it has *not* yet concluded. Moreover, this can be performed quickly, since it involves little more than a lookup of the current knowledge base.

Rules 3 and 4 illustrate one way in which an agent clock can be updated and also how direct contradictands can be kept from lingering, while other wffs may remain in the KB to facilitate further reasoning. Note also that, although this does "dismiss" the contradictands from further inferences, the "conflict-recognition" rule to be given below in 5, ensures that a record is kept in the KB of the former presence of a contradiction. Preserving this "historical" information is important in order to attempt a more solid repair of the contradiction.

As mentioned above, traditional formalisms, including most modal, temporal and nonmonotonic logics, suffer from the "swamping problem" (this is related to the "omniscience" problem of traditional logics of belief: all (infinitely-many) consequences of the axioms are theorems and hence are believed). As a result, in those logics, any possible clues as to how to proceed with reasoning when a contradiction is encountered are rendered ineffective by their own negations which are also derived from the contradiction. There have been some attempts to overcome the swamping problem, but so far only in the propositional case, and even so the essential time-dependency for real-time capabilities is still missing there.

Even though the problem of inconsistency is treated by some logics like paraconsistent logics, in reality most of the traditional logics do not note or repair inconsistencies, they just carry on with them. Nor in general do they provide for any special real-time status as needed by a real-world agent. On the other hand, active logics are intended for on-board use by an agent, not as an external specification of an agent.

In active logics, since the notion of inference is time-dependent, it follows that at any given time only those inferences that have actually been carried out so far can affect the present state of the agent's knowledge. As a result, even if directly contradictory wffs, P and ¬P, are in the agent's KB at time `t`, it need not be the case that those wffs have been used by time `t` to derive any other wff, Q. Indeed, it may be that `t` is the first moment at which both P and ¬P have simultaneously been in KB.

By endowing an active logic with a "conflict-recognition" inference rule such as that in 5, *direct* contradictions can be recognized as soon as they occur, and further reasoning can be initiated to repair the contradiction, or at least to adopt a strategy with respect to it, such as simply avoiding the use of either of the contradictands for the time being. The `Contra` predicate is a meta-predicate: it is about the course of reasoning itself (and yet is also part of that same evolving history).

$$(5)\quad \begin{array}{ll} \texttt{i} & : \quad \texttt{P, } \neg\texttt{P} \\ \texttt{i+1} & : \quad \texttt{Contra(i, P, } \neg\texttt{P)} \end{array}$$

The idea then is that, although an indirect contradiction may lurk undetected in the knowledge base, it may be sufficient for many purposes to deal only with direct contradictions. Sooner or later, if an indirect contradiction causes trouble, it may reveal itself in the form of

a direct contradiction. After all, a real agent has no choice but to reason only with whatever wffs it has been able to come up with *so far*, rather than with implicit but not yet performed inferences. Moreover, since consistency (i.e., the lack of direct or indirect contradictions) is, in general, undecidable, all agents with sufficiently expressive languages will be forced to make do with a hit-or-miss approach to contradiction detection. The best that can be hoped for, then, seems to be an ability to reason effectively in the presence of contradictions, taking action with respect to them only when they become revealed in the course of inference (which itself might be directed toward finding contradictions, to be sure).

Unlike most NMR formalisms, we do not attempt to capture the (usually undecidable) absolute truth about what is consistent with what is known; this is in general impossible for real agents. If nothing is *already* known that would prevent a default conclusion, then the agent has little choice except to draw that conclusion, and this is what an active logic does. If later (with more time) the agent discovers a consequence of its beliefs that in fact should have prevented that conclusion, it is only at that later time that it can be withdrawn, and this is what active logic makes possible. In principle, in the limit, active logic should, in special cases at least, provide the same default conclusions as standard NMR formalisms; this is a topic of current investigation.

Several example problems were solved this way in real time. For instance, during the planning, new information could become available in contradiction with existing beliefs. In that work, contradictions were treated in conjunction with default rules, where the rule that "P follows by default from Q" can be represented as in 6:

$$(6)\quad \begin{array}{ll} \texttt{i} & : \\ \texttt{i+1} & : \end{array} \frac{\texttt{Q, -Know(-P,i), Now(i)}}{\texttt{P}}$$

Thus if ¬P is not known at the current time, and if Q is known, then P is inferred by default at the next time step. However, it may turn out that at a later time, evidence for ¬P becomes known and a contradiction results. In the past work the particular example problems allowed for a very simple expedient in such cases: disinherit the default conclusion and accept the non-default evidence.

But while disinheriting contradictands is a reasonable first step, it is often not enough even to "defuse" the contradiction for long. P and ¬P may have come into KB for reasons that are still in force and the system may re-derive P and ¬P, or other similar conflicts, later on. Thus, in [36, 25, 47] we have investigated ways to allow an active logic-based reasoner to retrace its history of inferences, examine what led to the contradiction, and perform metareasoning concerning which of these warrants continued belief.

However, in general such an expedient is far too naive to be useful, and instead more sophisticated conflict-resolution methods are needed. Current research is aimed at the development of a *typology* of contradictions, which will allow appropriately specific methods to be applied to individual cases [5]. This in turn will provide for much more useful real-time deadline planning in which new evidence can be weighed against old along multiple dimensions.

# 5 Active Logic and Time-situated Agency

Our approach to implementing agency involves three conceptual "planks", building upon each other. First, there is the issue of representing aspects of the environment (keeping in mind that we have detailed three sorts of environment which are relevant to agency, and which must be tracked by an agent: temporal, physical and social). For many object-level behaviors, it is not necessary to have an explicit representation of the processes that the system performs, other than the programming or mechanism that produces the behavior at appropriate times. This thought guides a great deal of current work on physically integrated autonomous agents (see [15, 44] for an overview). However, in light of the complexities of social integration, and especially linguistic interaction, as well as the need to engage in meta-behavior, such as dialog about dialog, or reasoning about one's intentions, we believe that rich knowledge representations are needed, to support robust and controlled reasoning with and about that knowledge. In the case of meta-dialog, for instance, we are influenced by work such as [28, 29, 50, 46, 35], that proposes detailed logical representations of a range of dialog phenomena.

The second plank of our approach is an ability to effectively *use* such representations in inference to be able to notice interesting phenomena, such as implications of what has been done, recognizing resulting anomalies, and deciding what can be done about it. For use in a real-time agent, this (meta-)reasoning can not be off-line, but must be integrated within the normal reasoning behavior of such a system. This is, first, because the information which would need to be brought to bear in, for instance, adjudicating a contradiction is precisely that information contained in the current knowledge base; second, because the world does not stop changing, nor the agent stop gathering information while meta-reasoning occurs, and this information may well be relevant not just to the current problem being reasoned about, but also to the larger goals in the service of which the reasoning is being pursued (for instance, changes in the world may dictate abandoning a certain project altogether and beginning something else); finally this is required in light of the high worst-case complexities of reasoning with such rich representations and meta-level phenomena: one would not want to calculate all possible inferences before proceeding to act. A truly well integrated reasoner would not only continue to attend to its reasoning in light of the world, but continue to attend to the world in light of the progress of its reasoning, making decisions and adjustments as seemed appropriate.

This leads to the third plank, integration of reasoning with acting and non-logical processing. It is important not just to be able to observe and reason about those observations, but also be able to act—whether by giving a command to a domain controller, or initiating a speech-act. It is not enough to deduce that such an action has occurred, or should occur, or even to adopt an intention to cause it to happen. Reasoning about anomalies and meta-dialog is not enough. To be effective such reasoning should be integrated with normal functioning, being able to affect object level processes and observe the effects of these processes.

Active logics provide an integrated framework for these three conceptual planks. Aspects of the environment are represented as first order formulas in the knowledge base. Such formulas might represent perceptions of a user's utterance, observations about the state of

the domain, or rules added by a system administrator. Inference rules provide the mechanism for "using" the knowledge for reasoning. One aspect of active logic especially important in the current context is its robust ability to continue to reason normally as formulas are added, changed or deleted from its knowledge base. In other words, the evolving knowledge base is naturally integrated into the ongoing reasoning processes. We extend the logic with one special proposition, *call*, which, if it is ever proved, will initiate external action (that can be reasoned about and tracked through observation).

This makes active logic a good candidate for a reasoning agent which is expected to observe and interact in real time with a continually changing world or domain. Our current implementation of active logic, as represented in ALMA [48], already has this ability to initiate, observe and respond to external events and non-logical processes; [39] outlines planned extensions to this ability.

# 6 Accomplishments

Active logics have been successfully applied, in one form or another, to all the requirements of agency listed in section 2, and has been designed to meet the desiderata suggested in 1.

For instance, given the mechanisms outlined, it becomes possible in AL to explicitly encode time-sensitive expectations (in 30 minutes I should be at the airport). When the time arrives, the expectation can be transformed to an explicit belief (I am at the airport). Now this belief may well contradict something currently believed (I am *not* at the airport, but stuck in traffic), and this will trigger the appropriate adjudication mechanisms. Most important for the current discussion, this will involve discarding all derivations which relied on that assumption. Thus, if my expectation that I would be at the airport (now) was a premise used in further reasoning and planning (intention formation), those derivations will have to be discarded and re-considered.

In [19, 38, 33, 37] active logics were used to investigate such planning in deadline situations, where the planning process itself must take into account how long that very planning is taking, so that the plan can be completed and executed before a deadline is past. Rule 7 illustrates a simple rule that keeps track of time remaining to a deadline. More sophisticated rules were used to estimate how long further plan refinement will take, and to thus abandon plans that are expected to take too long to complete and execute.

$$(7)\quad \frac{\texttt{i} \;:\; \texttt{Deadline(d), Now(i)}}{\texttt{i+1} \;:\; \texttt{Remaining-time(d-i)}}$$

In [48] it has been shown how AL can provide a uniform platform for reasoning and plan execution in the context of an agent replanning its action when the plan it was enacting failed. Since the agent's original plan failed while it was executing the plan, the agent cannot backup to a preceding state, instead it has to try to accomplish its goal from its current state. In that work, a plan could contain both external actions to be performed and

goals to be met. A plan is taken for execution if its postcondition implies the agent's current goal, the preconditions are met and the goal is not already true. If a precondition of a plan is not met, then that precondition is made a goal.

Active logics have also been applied to reasoning about others' ongoing reasoning [17, 18], and to dialog, especially in pragmatic inferences that support dialog [24, 41] and in representation of meta and mixed-initiative dialog [43, 1, 49].

# 7 Future Directions

Although the work that has been done to date demonstrates that active logics have the desired flexibility for real-world reasoning applications, effective symbolic reasoning requires grounded symbols [26, 27, 2]. Planned research outlined in [39] envisions a physically, temporally and socially embodied agent which can generate symbols from sensor data, and then learn to reason with those symbols to support and enhance behavioral competence. In the most general sense, the proposed research is meant to cast light on how reasoning agents model their environment, establish (provisional) policies for thinking and acting, and, more importantly, (learn to) modify those models and policies in light of ongoing experience—including observation of, and advice from, other (human and artificial) agents.

We are also moving forward with more sophisticated contradiction-handling mechanisms [5], and are developing a dialog agent which makes extensive use of metareasoning and metadialog, including the ability to negotiate the distinction between the use and mention of a word [3], in order to support improved conversational adequacy [42].

In this essay we have suggested that a realistic implementation of agency will not be possible without simultaneously considering the various aspects of embodiment and their interrelations. We discussed in particular the physical, social and temporal aspects of embodiment, and detailed some important requirements of agency—continual observation, contradiction handling, introspection, and meta-reasoning—which emerged from these reflections. Finally we showed how attention to temporal embodiment can help one build a reasoner capable of meeting the various desiderata of real-world agency.

# References


[1] C. Andersen, D. Traum, K. Purang, D. Purushothaman, and D. Perlis. Mixed initiative dialogue and intelligence via active logic,. In *Proceedings of the AAAI'99 Workshop on Mixed-Initiative Intelligence*, 1999.

[2] Michael L. Anderson. Embodied cognition: A field guide. *Artificial Intelligence*, forthcoming.

[3] Michael L. Anderson, Yoshi Okamoto, and Don Perlis. The use-mention distinction and its importance to HCI. In *Proceedings of the Sixth Workshop on the Semantics and Pragmatics of Dialog*, 2002.

[4] Michael L. Anderson and Don Perlis. Symbol systems. *Encyclopedia of Cognitive Science*, 2002.

[5] Michael L. Anderson and Don Perlis. Detecting, classifying, and handling contradictions in a large, dynamic information environment, *forthcoming*. University of Maryland, College Park.

[6] Elizabeth Anscombe. *Intention, 2ed.* Harvard University Press, Cambridge, MA, 1963.

[7] Manjit Bhatia, Paul Chi, Waiyian Chong, Darsana P. Josyula, Michael O'Donovan-Anderson, Yoshi Okamoto, Don Perlis, and K. Purang. Handling uncertainty with active logic. In *Proceedings, AAAI Fall Symposium on Uncertainty in Computation*, 2001.

[8] Michael Bratman. *Faces of Intention.* Cambridge University Press, Cambridge, UK, 1999.

[9] Susan E. Brennan. The grounding problem in conversations with and through computers. In S.R. Fussell and R.J. Kreuz, editors, *Social and Cognitive Psychological Approaches to Interpersonal Communication*, pages 201–225. Lawrence Erlbaum, 1998.

[10] Susan E. Brennan. Processes that shape conversation and their implications for computational linguistics. In *Proceedings of the 38th Annual Meeting of the Association for Computational Lingusitics*, 2000.

[11] Susan E. Brennan and Eric A. Hulteen. Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems*, 8:143–151, 1995.

[12] F. Brown, editor. *The Frame Problem in Artificial Intelligence.* Morgan Kaufmann, 1987.

[13] Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Chuchill. *Embodied Conversational Agents.* MIT Press, Cambridge, MA, 2000.

[14] Waiyian Chong, Michael O'Donovan-Anderson, Yoshi Okamoto, and Don Perlis. Seven days in the life of a robotic agent. In *Proceedings of the GSFC/JPL Workshop on Radical Agent Concepts*, 2002.

[15] Andy Clark. *Being There.* MIT Press, Cambridge, MA, 1996.

[16] Donald Davidson. *Essays on Actions and Events.* Oxford University Press, Oxford, UK, 1980.

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

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

[19] J. Elgot-Drapkin. Using step-logic for tractable non-monotonic inference-researchsummary. In *Working Notes of the AAAI Workshop on Tractable Reasoning*, San Jose, CA, 1992.

[20] J. Elgot-Drapkin, S. Kraus, M. Miller, M. Nirkhe, and D. Perlis. Active logics: A unified formal approach to episodic reasoning. Technical Report UMIACS TR # 99-65, CS-TR # 4072, Univ of Maryland, UMIACS and CSD, 1993.

[21] J. Elgot-Drapkin and D. Perlis. Reasoning situated in time I: Basic concepts. *Journal of Experimental and Theoretical Artificial Intelligence*, 2(1):75–98, 1990.

[22] Harry Frankfurt. *The Importance of What We Care About.* Cambridge University Press, Cambridge, UK, 1988.

[23] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press, Cambridge, MA, 1988.

[24] J. Gurney, D. Perlis, and K. Purang. Active logic and Heim's rule for updating discourse context. In *IJCAI 95 Workshop on Context in Natural Language*, 1995.

[25] J. Gurney, D. Perlis, and K. Purang. Interpreting presuppositions using active logic: From contexts to utterances. *Computational Intelligence*, 1997.

[26] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

[27] Stevan Harnad. Problems, problems: The frame problem as a symptom of the symbol grounding problem. *Psycholoquy*, 4(34), 1993.

[28] Jerry Hobbs. Ontological promiscuity. In *Proceedings ACL-85*, pages 61–69, 1985.

[29] C. H. Hwang and L. K. Schubert. Episodic Logic: A comprehensive, natural representation for language understanding. *Minds and Machines*, 3:381–419, 1993.

[30] Haythem O. Ismail and Stuart C. Shapiro. Two problems with reasoning and acting in time. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Seventh International Conference*, 2000.

[31] Emiel Krahmer, Marc Swerts, Mariet Theune, and Mieke Weegels. Error detection in spoken human-machine interaction. In *Proceedings of Eurospeech'99*, Budapest, Hungary, 1999.

[32] Emiel Krahmer, Marc Swerts, Mariet Theune, and Mieke Weegels. Problem spotting in human-machine interaction. In *Proceedings of Eurospeech'99*, Budapest, Hungary, 1999.

[33] S. Kraus, M. Nirkhe, and D. Perlis. Deadline-coupled real-time planning. In *Proceedings of 1990 DARPA workshop on Innovative Approaches to Planning, Scheduling and Control*, pages 100–108, San Diego, CA, 1990.

[34] J. McCarthy and P. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence*, pages 463–502. Edinburgh University Press, 1969.

[35] Susan W. McRoy, Susan Haller, and Syed Ali. Uniform knowledge representation for language processing in the b2 system. *Journal of Natural Language Engineering*, 3(2/3):123–145, 1997.

[36] M. Miller and D. Perlis. Presentations and this and that: logic in action. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, Boulder, Colorado, 1993.

[37] M. Nirkhe. *Time-situated reasoning within tight deadlines and realistic space and computation bounds.* PhD thesis, Department of Computer Science, University of Maryland, College Park, Maryland, 1994.

[38] M. Nirkhe, S. Kraus, M. Miller, and D. Perlis. How to (plan to) meet a deadline between *now* and *then*. *Journal of logic computation*, 7(1):109–156, 1997.

[39] Tim Oates, Michael L. Anderson, and Don Perlis. Learning to reason with grounded symbols, *forthcoming.* University of Maryland, College Park.

[40] Tim Paek and Eric Horvitz. Uncertainty, utility and misunderstanding: A decision-theoretic perspective on grounding in conversational systems. In *Proceedings, AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, 1999.

[41] D. Perlis, J. Gurney, and K. Purang. Active logic applied to cancellation of Gricean implicature. In *Working notes, AAAI 96 Spring Symposium on Computational Implicature.* AAAI, 1996.

[42] D. Perlis, K. Purang, and C. Andersen. Conversational adequacy: mistakes are the essence. *Int. J. Human-Computer Studies*, 48:553–575, 1998.

[43] D. Perlis, K. Purang, D. Purushothaman, C. Andersen, and D. Traum. Modeling time and meta-reasoning in dialogue via active logic. In *Working notes of AAAI Fall Symposium on Psychological Models of Communication*, 1999.

[44] Rolf Pfeifer and Christian Scheier. *Understanding Intelligence.* MIT Press, Cambridge, MA, 1999.

[45] R. Picard. *Affective Computing.* MIT Press, Cambridge, MA, 1998.

[46] Massimo Poesio and David R. Traum. Conversational actions and discourse situations. *Computational Intelligence*, 13(3), 1997.

[47] K. Purang. *Systems that detect and repair their own mistakes*. PhD thesis, Department of Computer Science, University of Maryland, College Park, Maryland, 2001.

[48] K. Purang, D. Purushothaman, D. Traum, C. Andersen, D. Traum, and D. Perlis. Practical reasoning and plan execution with active logic. In *Proceedings of the IJCAI'99 Workshop on Practical Reasoning and Rationality*, 1999.

[49] David Traum, Carl Andersen, Yuan Chong, Darsana Josyula, Michael O'Donovan-Anderson, Yoshi Okamoto, Khemdut Purang, and Don Perlis. Representations of dialogue state for domain and task independent meta-dialogue. *Electronic Transactions on Artificial Intelligence*, forthcoming.

[50] David R. Traum, L. K. Schubert, M. Poesio, N. G. Martin, M. Light, C. H. Hwang, P. Heeman, G. Ferguson, and J. F. Allen. Knowledge representation in the TRAINS-93 conversation system. *International Journal of Expert Systems*, 9(1):173–223, 1996.