On Planning: What is to be Done?

Bruce Donald

87-890
December 1987

Department of Computer Science
Cornell University
Ithaca, NY 14853-7501

On Planning: What is to be Done?

Bruce R. Donald

Computer Science Department
Cornell University
Ithaca, NY 14853

1. Introduction

Domain-independent planning with a STRIPS-like action model [STRIPS] has become a "classical" problem in Artificial Intelligence. In some sense David Chapman [Z] has had the final word on classical planning, by providing a precise algorithm together with correctness, completeness, and complexity theorems. In light of his results one can imagine the planning unit of undergraduate AI courses being completely rewritten. The situation is analogous to the period when it was realized that heuristic search was algorithmic; like search, one can now expect classical planning to be adopted as a prodigal son of computer science proper.

As Chapman points out, however, the action representation assumed for his results is quite weak: it is unrealistic for real-world problems, and many domains cannot be formalized in it. Therefore, extensions to the power of the action representation are desirable. Such extensions are the focus of [A, AK]. These papers are more speculative, and less precise than [Z]; the development is informal, and thematically they seem closely allied with Naive Physics [Hayes]. Unfortunately, negative results of [Z] suggest that such extensions probably make even plan verification $\mathcal{NP}$-hard.

In this paper I will discuss the results of [Z] and the extensions proposed by [A, AK] , highlighting what I believe to be their main points and contributions.

The most interesting material comes later, when I explore some issues raised by but not settled in the papers. That section discusses what is to be done in planning, and in particular, describes:

- A modal truth criterion for the planner of [AK] .
- A "reduction" of planning in time to classical planning and a discussion of the relative power of the two methods.
- A formal framework towards describing planners with partial deductive closure (derived side effects).
- A discussion of failure analysis, error diagnosis, and recovery in domain-independent planning.
- An exploration of the mathematical model of time (and space) proposed by Allen.
- A discussion of the relative difficulty of classical planning, versus planning with a dense model of time.

2. Planning from [A] to [Z]: Description of the Papers

I will now discuss the results of [Z] and the extensions proposed by [A, AK] . I will assume some basic knowledge of classical planning, and so will not spend a lot of time reviewing what a plan is. the STRIPS action representation and so forth. This review section outlines what I believe to be the main points and contributions of the three papers. However, the details and remainder of the paper assume greater familiarity than available from the review. For example, I will not reproduce the details of clear and well-written ideas such as Chapman's modal truth criterion. However, I will restate and explain certain scruffier, less precise descriptions, such as the operation of the planner in [AK].

Interspersed in this section is a critique and analysis of the papers. In concession to readability, I have not always taken explicit credit for new ideas; you'd have to read the papers to notice this.

2.1. Chapman [Z]

[Z] assumes a restricted version of the classical planning paradigm. The STRIPS action model is assumed; all pre- and post-conditions in a plan step must be atomic propositions (possibly with variables). No derived side-effects, conditional actions, or functional dependency of the effects on the input situation are allowed. An output situation is created from the input situation by deleting the denied propositions and adding the asserted propositions to the input situation. The goal and input situation are represented as a conjunction of propositions. A non-linear plan is viewed as a set of partially-ordered plan steps. The planning problem is to produce a non-linear plan for the conjunctive goals.

Under these assumptions, Chapman is able to derive a *modal truth criterion* for determining when a given proposition is necessarily true in a given situation. This criterion is at the heart of the paper. Since the quantification in the modal truth criterion takes place over known, finite sets, it may be interpreted as a a non-deterministic procedure. As such it provides an algorithm for fast (linear-time)[1] plan verification. When incorporated into a breadth-first search, the criterion yields a semi-decision procedure for plan construction which is complete (up to decidability), correct, and produces an optimal[2] plan if one exists. A semi-decision procedure is the best we can do, since even with this restricted action representation, planning is equivalent to the halting problem.[3]

This is not all. Not only is TWEAK the first complete planner, but the modal truth criterion may be employed to analyze all previous classical planners. In particular, it provides a precise description of plan modification operators (such as

[1] In the number of steps in the plan.

[2] Shortest length in steps.

[3] While the proof requires that the input situation is allowed to be infinite, there is reason to believe that planning is undecidable even with finite initial state.

promotion and demotion) and codesignation constraints (introduced by [Stefik]). Anyone who has tried to slog through early planning papers such as [Tate] will immediately recognize the value of this contribution.

Chapman discusses extensions to the action representation of TWEAK, raising the issues of overlapping actions, actions with duration (i.e., that are not instantaneous), and actions with derived side-effects. These issues are the focus of [A, AK] , and are also discussed below. Is there an extension of TWEAK with a more powerful action representation?

Unfortunately, the modal truth criterion and its corollaries depend critically on the weakness of the action representation, and the discreteness and non-denseness of time. It also relies on the existence of an infinity of constants; this is because allowing codesignation restrictions to a finite set makes plan verification $\mathcal{NP}$-hard. For example, the blocks world with a finite table cannot be formalized in TWEAK's action framework. Chapman acknowledges this, but points out certain unpleasant effects of extending the action representation: If we allow functional dependence of postconditions on preconditions, planning becomes undecidable.[4] Furthermore, if functional dependence, conditional actions, or derived side-effects are allowed, then plan verification becomes $\mathcal{NP}$-hard.

However, it is not clear *a fortiori* that the extensions proposed by [A, AK] lead to a necessarily intractable planning problem. The discussion I will present, however, jibes with Chapman's conjecture that they must.

Chapman's thesis is amazingly concise and well-written. This makes it hard to summarize without quoting; the results are quite deep and deserve more discussion. Let us postpone this discussion until the second part of this paper: there I will try to apply some of Chapman's methods to the problems of [A, AK] .

2.2. Allen and Koomen [AK]

There is overlap between [AK] and [A] ; when discussing an idea in their intersection, I will attribute it to the common author. By in large, in commenting on both [AK] and [A] , I will focus on planning-related issues and attempt to evaluate the representations in light of their impact on planning. Unfairly perhaps, this is at the expense of many other interesting side-issues, such as natural language evidence, belief structure, intentionality, and so on. We will, however, delve somewhat into naive physics connections with [AK,A].

Allen proposes extending the STRIPS action representation of classical planning. In particular: Allen's actions have finite duration; they may overlap (in any way), and they may interact synergistically to enable or prevent goal attainment.

[4] Even with finite initial state.

Essentially, this amounts to introducing a model of time, and indexing the propositions in the pre- and post-conditions of actions by this model. Thus, for example, instead of the precondition Clear(x) in the blocks world, we would now have

$$\text{Clear}(x) \text{ holds over } T_x,$$

where T_x is a time interval.

To model time, Allen introduces a non-numerical axiomatization of abstract *time intervals*, and a set of binary predicates on time intervals such as Meets, Before, After, Starts, Finishes, Overlaps, etc. These relations, which have the form

$$T_x \text{ Before } T_y,$$

also appear in the pre- and post-conditions of an action. It is critical to the representation that two intervals are allowed to Meet, that is, to touch without overlapping. [Allen and Hayes] show that in fact all the other predicates can be expressed in terms of existentially quantified conjunctions of Meet propositions. The model of time raises interesting mathematical questions; for example, the open (or closed) intervals on the real line do not satisfy the Meet axioms. In a later section, we will attempt to construct a continuous model of time which does satisfy them. The task proves to be quite subtle.

Associated with the Meet axioms are certain deduction rules. A planner using Allen's extended action representation appears to require deductive closure of these rules. To this end, Allen proposes a constraint propagation algorithm. For example if an effect of an action is T_1 During T_2, then if T_2 Meets T_3 is known to be true, these propositions can combine to assert

$$T_1 \text{ Before } T_3. \qquad (*)$$

We must be able to compute derived side effects in this case, for otherwise it would be impossible to know whether the preconditions for subsequent actions—eg., $(*)$—were satisfied. Note that this alone does not imply that such a planner must do exponential work; there could, *a priori*, exist constraints in the domain of temporal inference that allow a fast deductive procedure. Unfortunately, such a procedure probably does not exist: determining the consistency of a set of interval relations is $\mathcal{NP}$-complete [Kautz and Vilain].[5]

Allen introduces an interesting notion of causality in [A] ; in the restricted domain of [AK] , he detects *causal gaps* when an assertion—eg., a goal or subgoal—is not the effect of some action in the plan. The planning problem is viewed as

[5] Determining the consistency of a set of constraints is the same as determining whether one relation follows from another. Kautz and Vilain also show that Allen's constraint propagation algorithm is incomplete.

eliminating these causal gaps by introducing plan steps and temporal constraints. An example of a temporal constraint is "A must be Before B" for the plan to work.

Allen proposes a planner based on this idea. Step one of the planner entails examining all time intervals in the plan. If identifying two (or more) time intervals is legal, and eliminates a causal gap, then the identification is performed. In the worst case, however, this requires iterating over the powerset of all time intervals. Thus we conclude that even if computing the deductive closure takes constant time, the planner of [AK] is exponential. As suggested by the result of [Kautz and Vilain], and my modal truth analysis below, this is probably necessarily so.

Step two of the planner is so vague that it is impossible to tell whether it is correct or complete. The point is somewhat moot, since later we will explore a correct and complete, albeit exponential-time algorithm.

This paper raises many interesting questions. Does there exists a modal truth criterion, à la Chapman, for Allen's action representation? Is planning with time "more powerful" than classical planning? How can some form of deductive closure be used in a planner? What are the implications of Allen's mathematical model of time? Partial answers to these questions are described later in this report.

2.3. Allen [A]

In [A] , Allen proposes an ambitious model of time, events, processes, actions, causality, and intentionality. The model is based on first-order predicate calculus and has close ties to naive physics [Hayes]. In some sense this paper is the "theory" behind [AK]; like naive physics, it has computational effectiveness as an eventual goal, but this aspect is not emphasized. The stress is instead on the representation, and its logical consequences. The portion of the theory which is effective (or nearly so) is the planning submodule described in [AK] ; much of [A] only glancingly addresses planning.

Allen's model is almost entirely descriptive. At a high level, one might view it as follows. In certain fields there are developments we might call *enabling technologies*. For example, in mathematics there is category theory; in software engineering there is EMACS. These technologies do not directly allow one to "do" more, but rather allow one to think more clearly about the problem (and hence get more done). [A] is reminiscent of such developments, in that it is in essence a representation without a task; whether it is useful or not, however, remains to be shown. Personally, I have some reservations about such work; but let us proceed.

In describing his representation, I'll try to emphasize connections to planning, and to naive physics. In relation to naive physics, we will be primarily interested in Allen's model of time and space.

Suppose we view classical planning as follows. Let X denote the set of all possible propositions. A goal then is a conjunction of elements in this set. Allen introduces a set of objects T called time intervals, which are characterized by the axioms we reviewed above. We may view his model as a theory, in the sense that it is the deductive closure of these axioms. Formally, he extends X to $X \times T$; that is, he indexes the propositions in X by the time intervals. Thus $(p, t) \in X \times T$ denotes the proposition p holding over time interval t.[6]

Some Thoughts on Time

Since Allen's representation is proposed as a model for problem-solving and reasoning in general, it seems fair to discuss it in this more general context.

Time may be viewed as discrete; this is essentially the STRIPS model, which views time as the integers. If time is discrete, then the Meet axioms are satisfied by sets of the form $[a, b]$, which denotes all integers between a and b, inclusive.

However, a continuous[7] representation of time is required for Allen's action representation, and for other applications such as Qualitative Process (QP) theory [Forbus]. The reason for the first is introspective: Allen desires to represent events which have duration, and which take place over continuous intervals. The natural representation for continuous time is intervals on the real line.

QP theory and other qualitative reasoning work (eg., [Williams]) rely on a continuous model of time for several deep reasons:

- In the literature, processes are associated with continuous functions. Qualitative reasoning requires reasoning about the continuous functions whose domain is time. In particular, the reasoning is restricted to this category. If time is discrete,[8] then *all* functions are continuous. This is clearly undesirable.
- Limit analysis [Forbus, Williams] depends crucially on the differential structure of the real line. Discrete time has a trivial differential structure, and hence is inadequate.
- Limit analysis and QP theory depend critically on the intermediate value theorem. This theorem assumes that the time intervals are connected. For discrete time, this is of course, false.

[6] Actually, to be completely general, we really should form the product of X and the powerset of T.

[7] For a long time the use of "continuous," as applied to a space, confused me. In modern mathematics it is only applied to functions. What is meant, I think, is that time is a linear continuum. In classical differential geometry [Riemann], a manifold was called a *continuous* space. I'll use "continuous" in this sense.

[8] Formally, by "discrete", we mean having the discrete topology. For the integers, it is the same as the order and subspace topologies.

In aggregation research, when [Weld] tried generalize the intermediate value theorem while relaxing the connectedness assumption, he found that the only functions he could aggregate were surjective linear functions, of the form $f(t) = t + b$.

I believe this makes a strong argument for finding a continuous model of time. I'll attempt to construct one later. For now, suffice it to say that:

(1) The open (or closed) intervals in the real line do not work, since they cannot Meet.

(2) There is a view in naive physics that "points are bad." The interval-based logics come as a rebellion to logics which started with truth at points (in the real line), and attempted to extend to intervals. Such logics would start with propositions like Holds(run,p), meaning "at time point p, Sam was running;" to extend running to an interval, one quantifies over all points in the interval. The problem is that it is impossible to isolate what the *run* proposition should be at an instant; running cannot be decomposed easily. In short, many events, such as running, were felt to have the most natural definitions over *intervals*, not at points.

A time logic without points also has the following introspective justification. If a man gets married, at the instant of marriage, is he married or not? Time seems ill-conditioned at this point. In physical situations (eg., boiling), the behavior of the physical system is often ill-conditioned at the transition point. It would be convenient to have the two intervals Meet, without worrying what happens at the endpoint. Perhaps one should not be able to refer to the endpoint at all; this certainly seems the view of [Allen and Hayes]. This makes it tricky to construct a consistent continuous model of time. As we saw from the analysis of discrete time, the logical consequences of the model impact the effectiveness of QP theory and other qualitative reasoning.

I have alluded to "Allen's model of space" as well as time. This is actually a taradiddle; he never explicitly mentions space. However, in the naive physics literature, [Hayes, Forbus, Fleck, Weld, Schmolze] there is the belief that two physical objects should be allowed to Meet, that is, to touch without overlapping, with no "space" intervening. This axiom is believed by naive physicists to be rather crucial. For example, the argument is made [Hayes, Fleck] that if a cup is on a table, then the two objects are perceived as Meeting with no space intervening. Modulo its order-dependent properties, Allen's time logic is also applicable to modelling space, since it contains the Meet axioms. Finding a continuous model of space which satisfies these axioms turns out to be subtle however; see [Forbus, Weld, Fleck]. I'll continue this discussion and describe such a model later on.

Imperfection and Perfection

Since Allen is interested in actions with duration, he makes a distinction between *events*, which are essentially atomic and take a "minimal" amount of time, and *processes* which take up finite time, and occur during "most" subintervals. All his definitions bifurcate along this distinction, which I will term the difference between *perfect* and *imperfect* events. This allows Allen to introduce axioms which allow deductions such as

"I ate the cake" implies *"I was eating it."*

while the converse need not hold. A more sophisticated model of causality than [AK] is introduced, along with some primitives for representing beliefs of agents about plans. This yields a model of intentional action.

The formalization is obviously exploratory, so one wishes to be gentle with it. In this vein, I have the following criticisms:

(1) Perfect occurrences ("events") are defined to occur over "minimal" time intervals. That is, if event E occurs over a time interval T, it occurs over no subinterval of T. That is, for all $T' \subset T$, E does not hold over T'. I cannot imagine such an axiom holding, unless it is by failure of the antecedent: Consider the real line. What kind of event takes finite time, but does not occur over subintervals? The axiom only makes sense for points, in which case the antecedent fails. I would have to be shown an example to believe this axiom is reasonable.

(2) Allen refers to a desire to exploit "grain-size" of time intervals, and to consider points as "very small time intervals." There is, in fact, a formal way to do this, called "Geometry without Points." We will discuss it later.

(3) Given the model of causality in [A] , there is a serious confusion over what a plan should be. We have the definition, "A plan is considered to be effective if the desired world is implied by the expected world." Allen gives a detailed example of one person's planning to hide a coat from another. In this example, however, we find that "...in his expected world Carole will see the coat, while in the desired world, she will not." I believe I see what the planner is supposed to so in the simplified domain of [AK] . It is hard to see in the more complicated domain with belief structure and intentionality. I'll focus on other issues in this paper, but it would be interesting to make this work.

(4) In this paper I will use the term *uncertainty* to refer to inaccurate, imprecise, or partial knowledge about the world. Analogously, we can talk about uncertainty of *control*, that is, in determining the effects of an action.[9] I believe it is appropriate to speculate on the robustness of algorithms and representations in the face of uncertainty. For example, it seems there would exist generalizations of [AK] to situations with uncertainty.

Allen alludes to the necessity of dealing with uncertainty in [A] . I believe, however, that his model of intentional action is brittle in the face of uncertainty.

[9] This does not imply a particular model of uncertainty, and in particular does not imply probabilistic uncertainty or "certainty factors."

For example, Allen discusses playing a C-chord on a piano. It is concluded that if playing the chord is intended, then playing the C note is also. But what if there is control uncertainty and a D is hit instead? Is playing the D intentional? What if, despite the control uncertainty, a C is still hit. Is this act intentional? What if an octave of the "desired" C is hit instead? What if a G-note is hit by mistake, but a C-chord still is played (although perhaps the wrong inversion)?

I do not view these situations as bizarre or contrived: the bug is that Allen views *every* action as intentional. Furthermore, he declares that for each action, there is a *unique* agent that caused it. The first assumption is too weak for dealing with uncertainty. The second is too weak for coping with synergistic interaction of agents. (For example, imagine two agents cooperating in playing the piano, or in a a simple assembly). Until these deficiencies are overcome, I think that [A] 's belief logic and intentional action structure will not be viable.

Finally, these examples bring out the importance of failure analysis in planning: how can we detect when plans have failed, and how can we recover from failures? I'll address this issue later.

Natural Stupidity

I will close with what I feel is a "deeper" criticism of Allen's representations.[10] I believe that they are overly syntactic, and that much of the semantic content is in fact conveyed in the English meaning of his primitives, not in the predicate calculus formalization. The basic approach of [A] is to introduce primitives such as BELIEVE, COMMITTED, INTEND, INTENTIONAL, CHANGE-MIND, and so on, and then to give them formal, syntactic properties. But part of the semantics is captured in the name; no one would claim the representation was perspicacious if the primitives were called things like G0034.

Let us take an example. Allen gives the following formalization of CHANGE-MIND, which roughly speaking, is supposed to be applicable to situations like "The professor had planned to pass the student, but changed his mind after reading the paper."

Necessary and Sufficient Conditions for
OCCUR(CHANGE-MIND(agent,property,ptime),time):
$\exists$ t1,t2 such that
MEETS(t1,time) and
MEETS(time,t2) and
BELIEVES(agent,property,ptime,t1) and

[10]Unfortunately, to be fair, it is no more true of Allen in particular than of a lot of AI in general.

BELIEVES(agent,not(property),ptime,t2)

I agree this formalizes some aspects—particularly the temporal interrelations—of an agent changing his mind. But it is not a "sufficient" characterization. Suppose $t_1 < ptime < t_2$, that is, $ptime \subset time$. The agent might have believed at t_1 that Holds(property,ptime) would be true (in the future). But the agent might not have liked the fact, and might have developed a plan during *time* to change things so that *property* would *not* hold over *ptime*. After the plan is successful, at time t_2, the agent believes that the property did not hold over *ptime*, because he made it not hold. This does not represent a change of mind during *time*, although it satisfies the formal prerequisites above.

The best kind of definitions would make sense even if the objects in them were called X, Y, and Z. In mathematics, for example, after seeing such a definition, it makes sense to call it, say, a "retraction;" but the English term is a convenience, and not itself definitional. When English nuances are still present in a formal scheme we end up with a no-man's land of kinda-sorta logic.

3. What is to be Done?

In this section, I will discuss several important planning issues raised by, but not settled in the three papers. I believe that these issues reflect central themes in domain-independent planning, and that they stimulate a research agenda for what is to be done next.

The highlights of this section are the following topics:

- Correctness and completeness for extended action representations. A modal truth criterion for the planner of [AK] is discussed.
- The relative power of planning in time versus classical planning. The relative power of classical planning versus algorithmic robot motion planning. The latter comparison is used to characterize the power of planning with a dense representation of time.
- An analysis of partial deductive closure mechanisms, for planning actions with derived side-effects.
- Failure analysis and execution monitoring in domain-independent planning.
- Mathematical models of time and space for planning and naive physics.

3.1. A Modal Truth Criterion for Planner of [AK]

We do not know whether Allen's planner is correct or complete. Since the extended action representation is desirable for planning in more realistic domains, let us develop a modal truth criterion for the action representation in [AK] . Our starting point naturally will be [Z] . The modal truth criterion can be interpreted as an exponential-time algorithm for plan verification. This is consistent with Chapman's conjecture that any extension of the TWEAK action representation makes the plan verification problem $\mathcal{NP}$-hard.

Here is the fundamental difference we must consider: if overlapping actions are permitted, they may interact synergistically. ([Z] made this observation). That is, two (or more) actions may conspire to assert a proposition that neither could assert alone. For example, if one action is hitting a nail with a hammer, and another is holding the nail (in the right place), then their penecontemporaneous interaction results in the nail being driven in (the right place). Symmetrically, actions may combine to prevent an effect: for example, inserting my thumb between the nail and the hammer prevents it from being driven in. Finally, prevention can also be conspiratorial; that is, a set of overlapping actions can conspire to prevent an effect that none could have prevented alone.

The existence of conspiracies necessitates quantification over the powerset of possibly overlapping actions. This is the fundamental reason why the modal truth criterion will be exponential. The generalization of establishing steps and white knights are establishing conspiracies; the generalization of clobberers are prevention conspiracies.

Formalizing the modal truth criterion : Conspiracies

Suppose we have a set of steps $\{ E_i \}$, and each E_i occurs over a time interval T_i. Denote this by $\mathcal{E} = \{ E_i, T_i \}$. Actions are represented as in [AK] . We wish to determine when $\mathcal{E}$ ***necessarily conspires*** to assert a proposition p to hold over some time interval t. We need a way to represent synergistic interactions, which we introduce as the ***conspiracy preconditions*** and ***conspiracy effects***. Conspiracies are represented as follows. Note that they are a generalization of the the "decompositions" of [AK]; I'll say more about this later.

Representation of Conspiracies

(1) Names of steps $\{ E_i \}$ occurring over $\{ T_i \}$. Each action has its usual precondition structure [AK] of propositions that must hold over time intervals, and relations that must hold among the time intervals.

(2) Names of possibly preventing steps $\{B_j\}$ occurring over $\{S_j\}$.[11]

(3) A set of codesignation relations among the variables in the preconditions of the $\{E_i\}$ and the $\{B_j\}$.

(4) A set of temporal relations among the $\{T_i\} \cup \{S_j\}$ and the time intervals in the preconditions of the $\{E_i\}$ and the $\{B_j\}$.

(5) CPX: The ***conspiracy preconditions***, which is a proposition combining (3) and (4) using disjunctions and conjunctions.

(6) CPE: The ***conspiracy effects.*** The effects (such as $\{q_k$ holds over $T'_k\}$) which are asserted if the individual preconditions of the $\{E_i\}$ are satisfied, and the CPX are satisfied. The CPE can override individual postconditions.

A proposition in the CPX, or in the individual preconditions of an action, has the form Holds(p, t) or tRs where R is a temporal relation. We say such a precondition is ***satisfied*** if it is necessarily true.

Now, note that a representation such as this is in fact necessary—in general we cannot infer the effects of synergistic interaction from the [AK] description of individual actions alone.

An Example

An example will help clarify this. Consider a variation of Chapman's *Gedankenplan* of trying to push one block of LEGGO onto another. We have:

E_1 occurs over T_1 where E_1 is "push down on the LEGGO block-x from the left side."

E_2 occurs over T_2 where E_2 is "push down on the LEGGO block-y from the right side."

B_1 is a possible preventor which might occur over time S_1. Its action is "put thumb under LEGGO block-z."

These actions have the obvious individual preconditions: for example, for E_1 to occur over T_1, a finger must be over block-x.

The conspiracy preconditions (CPX) are as follows:[12] (We ignore the uninteresting case where B_1 does not occur).

T_1 intersects T_2 $\wedge$
block-x $\approx$ block-y $\wedge$
(block-z $\not\approx$ block-x $\vee$

[11] To make this work, we may have to introduce a new time interval $\emptyset$ as a formal convenience. The assertion "E occurs over $\emptyset$" is equivalent to "the action E does not occur."

[12] $\approx$ denotes codesignation of variables.

$\exists$ time X, disjoint from S_1, such that $X \subseteq T_1 \cap T_2$).

Although I wrote this in structured English, it is easily expressible in Allen's time logic. The conspiracy effects, are of course, that the top block is pressed down onto the bottom block (holding over some time interval).

Note that a conspiracy is defined to contain its possibly preventing steps; constraints in the conspiracy keep the preventors where they can do no harm. Furthermore:

- As defined above, a conspiracy need not be contemporaneous. If we wish to enforce contemporaneity, we must add a ***temporal connectedness constraint*** that the union of the $\{T_i\}$ be connected.[13] Without this constraint, the establishing (or preventing) steps in the conspiracy may be separated in time. Thus it is possible to plan for the synergistic interaction of plan steps that do not overlap in time. As Chapman points out, this is often necessary, even in simple domains.
- The issue of temporal connectedness points out the importance of a continuous model of time: if the integers (or any discrete set) are used to model time, the only connected sets are points.

Now suppose that any action of some fixed type could overlap with any number of other instances of itself. Then we could not decide the synergistic interaction problem, since we would have to consider an infinite number of subsets of plan steps. For example, suppose there were only one action type: a molecule hitting a window. If a huge number of these actions can take place at the same time, then the state of the window could change. But we can never guarantee that it ***won't*** change unless we have a bound on the number of overlapping actions we must consider. To say this more formally, we want to make sure that, at worst, we quantify over all subsets of a ***finite*** set of actions.

For this reason, we will require that if n actions of the same type can interact, then the action must be entered n times into the set of all actions.[14] The modal truth criterion will define a procedure which can verify a proposition's necessary truth in time exponential in the size of this enlarged set.

This concludes the discussion of when a conspiracy $\mathcal{E}$ necessarily asserts a proposition p to hold over some time interval t.

The Generalized Modal Truth Criterion

Here is the modal truth criterion for Allen's planner. If t and s are time intervals, we will write $s \subseteq t$ iff

[13] Formally, this constraint is that both $K = \bigcup_i T_i$ and $K \cup \bigcup_j S_j$ are connected.

[14] With different indices, obviously.

$$s\ (= \text{during starts finishes})\ t.$$

Modal Truth Criterion for [AK] . *A proposition p necessarily holds over some time interval s if, and only if,*

(1) p is necessarily asserted to hold over some time interval t, where necessarily $s \subseteq t$.

Note that for (1) to hold, there must exist a conspiracy $\mathcal{E}$ which necessarily asserts p over t. That is, (1) is equivalent to

$$\exists \mathcal{E} \exists t \quad s \subseteq t \wedge \text{Asserts}(\mathcal{E}, \text{Holds}(p, t))$$

where $\mathcal{E}$ is a conspiracy, and Holds(p, t) codesignates with a proposition in its CPE. The number of conspiracies is bounded by 2^n where n is the size of the action set. Therefore determining the existence of such a conspiracy requires at worst exponential work, since it requires quantification over the powerset of all actions. Quantification over potential clobberers is guaranteed by the construction of the conspiracies' CPX.

The modal truth criterion is simple because the definition of conspiracies is complicated. It can be interpreted as a non-deterministic procedure:

(a) Choose an existing $\mathcal{E}$, or a new $\mathcal{E}$, or a mixed $\mathcal{E}$ as the establishing conspiracy. (A *mixed* $\mathcal{E}$ is a conspiracy formed from some existing plan steps, and some added steps).

(b) Choose a proposition u in the CPE of $\mathcal{E}$ such that u is asserted to hold over time t.[15]

(c) Make p and u codesignate.

(d) Choose and assert one of the following: $s = t$, $s \subset t$, s Starts t, or s Finishes t.

However, the most interesting plan modification operators are derived from the conspiracy preconditions (CPX). For example, the CPX for the LEGGO block example (above) can be viewed as a non-deterministic procedure also:

(a') Assert "T_1 intersects T_2."

(b') Assert "block-x $\approx$ block-y."

(c') *Either* assert "block-z $\not\approx$ block-x." *or* do all of the following:

Choose a time interval X.

Assert "$X \subseteq T_1 \cap T_2$."

Assert "X is disjoint from S_1."

Some of these steps involve more non-deterministic choices. For example the assertion "T_1 intersects T_2" can be accomplished by

[15] It is convenient to view individual steps as "conspiracies of one." Note conspiracies of one exist iff there are steps that are unclobberable by overlapping actions.

(a') Choose and assert one of the following: $T_1 = T_2$, T_1 Overlaps T_2, T_2 Overlaps T_1, T_1 Starts T_2, T_2 Starts T_1, T_1 Finishes T_2, T_2 Finishes T_1, T_1 During T_2, or T_2 During T_1.

It is likely that certain domains will have different kinds of CPX, and hence, different kinds of plan-modification operators. It would be interesting to explore this. It appears we cannot hope for a fast planner, however, if we want correctness and completeness for domains in which plan steps can interact synergistically.

As I've mentioned, this modal truth criterion also covers the case where steps separated in time can conspire together. Formally, this merely requires relaxing the temporal connectedness assumption on conspiracies. In practice, one would expect this to make the planner even less efficient.

Now, although the generalized modal truth criterion is potentially exponential in the number of plan steps, it is not necessarily exponential in the size of the *input*. In particular, if 2^n conspiracy representations are input to the planner, then the computation may no longer be exponential in the size of the input. Of course, this is not a necessary form of the input. The conspiracy representation can be constructed from a more compact input. An example of such an input would be the blocks world of HACKER [Sussman], where more than one block can fit on another. We will say more about inputting conspiracies in the next section.

[AK] employ a hierarchical decomposition of plan operators. For instance, MOVE is decomposed into PICKUP, HOLDING, and PUTDOWN. In our framework, these three steps denote the establishing conspiracy $\{ E_i \}$. To make MOVE a well-formed conspiracy, we would have to add all potentially overlapping preventors. If there are only two robot arms in this example, that means that there are also three possibly preventing steps, PICKUP$'$, HOLDING$'$, and PUTDOWN$'$; these are the $\{ B_j \}$. The CPX will keep the primed ($'$) actions from interfering with the unprimed actions. Thus MOVE is a conspiracy of size six.

3.2. On the Power of Planning In Time

Let us consider the planning problem and action representation addressed by [AK] ; call this problem "planning in time" to distinguish it from classical planning. Some obvious questions arise: Is planning in time somehow "more powerful" than classical planning? Can one be reduced to the other? What is the relative power of the methods?

Now, since [Z] has shown that classical planning is equivalent to the halting problem, we would expect that there exists some computation-theoretic reduction. Such a reduction is not particularly interesting since it does not elucidate the structure of the planning problems. In this section, I will sketch what seems a more perspicacious reduction of planning in time to classical planning. The reduction is exponential-time, but this doesn't matter if plan verification with time is exponentially hard anyway. What is most interesting are the kinds of plan operators introduced in the reduction.

Keep in mind which way the reduction is going; we want to see how a TWEAK-like planner might perform planning in time, if we give it some new operators. The development here is informal, since I don't want to obscure the idea by a lot of turgor.

A Sketch of the Reduction

First, we define one classical plan operator for each conspiracy. The representation for the operator is essentially the same as in the conspiracy representation; the preconditions are of the form

$$\mathrm{Holds}(p, t),$$

and

$$tRs$$

where t and s are time intervals, and R is a disjunction of conjunctions of temporal relations. We may have to define an exponential number of these operators. The preconditions for the classical plan operator must include both the CPX and the individual preconditions of the constituent subactions.

Next, we must introduce plan operators that permit temporal deduction. That is, we will use classical plan operators to derive the partial deductive closure of the temporal relations. Thus the resulting "plan" will contain, in addition to the usual kinds of actions, "actions" of the form "deduce and assert that T_1 overlaps T_2." These "actions" have the following form:
Example:

```
ASSERT-BEFORE(a,b,c)
        Preconditions:  a During b
                        b Meets  c
        Postconditions: a Before c.
```

To allow deductions such as

```
ODM(a,b,c)
        Preconditions:  a Meets b
                        b During c
        Postconditions: a Overlaps c   OR
```

```
a During c     OR
a Meets c
```

it is necessary to introduce a number of "convenience" operators such as

```
OR2(p1,p2)
        Preconditions:  p1
        Postconditions: p1 OR p2.
```

We will have to add a lot of temporal deduction operators. Also, in general we will of course need an infinite number of ORn operators; this is not an interesting point, since in any plan synthesis, only a finite number of them will be employed. Furthermore, these operators really represent a parameterized family of "isomorphic" operators.

At any rate, it is fairly clear that a TWEAK-like system can plan in time if we give it an exponentially large number of operators.

This reduction highlights two important questions about planning in time:

- How are synergistic interactions, in particular, conspiracies, to be computed and represented? There are a potentially exponential number of them; can we input a more compact representation and have them computed automatically?
- How should the temporal inference procedure be controlled?[16]

The reduction also shows that the representation of a planning-in-time problem is, in the best case, exponentially more compact than the representation of the same problem in a classical planning framework. In this sense, one could certainly make an argument that the representation captures the structure of the problem more concisely. I also feel that compared to classical planning, Allen's representation for actions is more natural, and can characterize many natural constraints in the problem in a more intuitive way.

The reason that I thought about the reduction, however, was because [AK] was fairly imprecise; I wanted to see whether a planner in time was really "more powerful" than a classical planner (as is claimed in the paper). Of course, it isn't—this is Chapman's result—but perhaps the representation is more convenient.

3.3. Deductive Closure

In this section we explore the issues of deductive closure and derived side effects. The deductive closure of a set of propositions is all the propositions that follow logically from the set. The deductive closure of the effects of a plan action is a set of derived side-effects; the postconditions are asserted, and are allowed to combine synergistically to assert new propositions. [Z] shows that in general, if derived side

[16] [AK] do address this issue to some extent, using hierarchies of *reference intervals*.

effects are allowed, then plan verification becomes $\mathcal{NP}$-hard. There are several reasons for exploring planners with deductive closure here, even if they are slow or incomplete:

- Derived side effects are desirable from an engineering point of view, since they represent a useful and powerful view of actions. For example, SIPE [Wilkins] is a planner which allows some derived side effects.
- Derived side effects are necessary to make Allen's time logic work in a planner.
- Deductive closure is probably necessary in any planner which uses the action, time, and belief representation in [A] .

Because giving a planner full deductive closure will in general result in intractability, I propose we examine the effects of allowing *partial* deductive closure. That is, the planner is allowed to compute some, but not all derived side effects of an action. A simple model of this might be something analogous to linear resolution:[17] "Create the output situation by deleting the denied propositions and adding the asserted propositions to the input situation. Then perform some small number of deductions by allowing the postconditions to interact with the other propositions in the output situation." Note that a truth maintenance system must be employed to ensure that when a proposition is deleted, all derived propositions that follow from it are deleted also.

It is possible to construct smarter algorithms for computing derived side effects; I don't particularly advocate the one above.

How far can we go with incomplete deductive closure? Let us explore the power of such partial deductive machinery. I'll propose a formal framework for deciding when plans can succeed in achieving the goal, even if the closure computation is incomplete.

We will examine the state space of all situations. The current situation will always be a point in this space. With complete knowledge of the world, and complete deductive closure, this point will always be known exactly. However, with only partial deductive closure, the situation can only be known to lie within some "region" in this space. We now formalize this notion.

Let X be the set of all possible atomic propositions (without variables). X includes both negated and non-negated propositions. A typical elements of X might be *(ON A B)*. Let $\mathcal{C}$ be the set of maximal, consistent subsets of X. That is, if $C \in \mathcal{C}$, then C is a subset of X such that for all p, $q \in C$, p and q are consistent. For example, suppose

$$X = \{ a, b, \neg a, \neg b \},$$

then

[17] This is an old idea; STRIPS did something like this.

$$C = \{ \{ a,b \}, \{ \neg a,b \}, \{ a,\neg b \}, \{ \neg a,\neg b \} \}.$$

We think of the proposition a as representing a *region* in C. The region is

$$\{ \{ a,b \}, \{ a,\neg b \} \}.$$

That is, a specifies the region in C where a is true. More formally, we view an element of C as a tuple

$$x = \{ x_\alpha \}_{\alpha \in A} \in C$$

for some index set A. x represents the assertion that

$$\bigwedge_{\alpha \in A} x_\alpha$$

is necessarily true. x represents a situation in the state space C.

Let Y be the set of non-negated propositions in X. We can view $x \in C$ as a set of consistent True and False assignments to Y. Therefore, x is a function

$$\begin{aligned} x : Y &\to \{ T, F \} \\ x_\alpha &\mapsto T \text{ or } F. \end{aligned}$$

which is natural, since C can be identified with a subset of 2^Y.

Now, suppose we have a region $R \subseteq C$. R represents the assertion that

$$\bigvee_{z \in R} z$$

is necessarily true. Conversely, any consistent proposition

$$\bigvee_\beta \bigwedge_\alpha p_{\alpha,\beta}$$

represents a region in C. These regions represent possible states of the world, in that they represent disjunctions of situations.

A good example of a region is the set of situations satisfying the preconditions of an action.[18] To determine the effects of actions we will introduce *forward projections.* [E] introduced forward projections in a geometric domain for robot planning; similar ideas in domain-independent planning are mentioned by [Rosenchein, Pednault]. Let ψ be an action, and let $x \in C$ be a situation in the precondition region for ψ. Then after executing ψ from situation x we obtain situation $x' \in C$. x' is a function of x and of ψ; we write this as

[18]For simplicity, we will consider partial deductive closure only for the classical plan representation, and not for Allen's planner.

$$x' = f_\psi(x).$$

f_ψ is called the ***exact forward projection of x under ψ***. It represents all propositions in X true after executing ψ from x, after we take the deductive closure.

Suppose, however, we only take a *partial* deductive closure. That is, we derive a few, but not all, of the derivable side effects of ψ. In this case the remaining, underived propositions are still true, but we cannot tell that they are true. In fact, the goal proposition(s) could even have been asserted, but the planner cannot necessarily know this. Let me stress this point: unless there is complete deductive closure, *reaching* the goal does not guarantee *recognizing* that it has been reached. In order for a plan to be guaranteed, the planner must recognize that the goal has been reached.[19] In short, the best possible characterization of x' is that it lies within some region of $\mathcal{C}$. Let us attempt to model this region. The question we want to answer is:

- Given a partial deductive mechanism, when can the planner guarantee that a plan will attain the goal? That is, when can the planner recognize that a plan will attain the goal?

First, we define the *approximate* forward projection of x under ψ, denoted $F_\psi(x)$, to be the region in $\mathcal{C}$ obtained as follows: $F_\psi(x)$ is defined by the conjunction of the propositions ψ asserts, and the negation of propositions ψ denies. Of course, how the variables in these propositions are bound depends on x. Thus,

$$F_\psi(x) = \Big(\bigwedge_{p \in Addlist(\psi)} p(x)\Big) \bigwedge \Big(\bigwedge_{q \in Deletelist(\psi)} \neg q(x)\Big).$$

$F_\psi(x)$ is a region in $\mathcal{C}$; it is in some sense an outer bound on where x' can be;[20] The reason the forward projection is a region is that it does not specify the validity of the propositions not mentioned in the effects of ψ. It is conservative when ψ is consistent, so we have

$$x' \in F_\psi(x).$$

Henceforth, all forward projections are assumed to be approximate.

Now, we introduce the following execution monitoring technique. In a "real" planner it is desirable to monitor the execution somehow. In this case, we will allow the execution system to check the value of some set of propositions after executing a plan step. Call this set x^*; note that it is reasonable to allow x^* to vary with the plan step being monitored. The execution monitor is allowed to check each proposition in x^* for validity; the existence of such a mechanism is justified by the

[19]This distinction between reachability and recognizability is a theme in recent robotics literature [LMT, E].

[20]To make this work, we must of course have that ψ's effects are consistent.

necessity of precondition checking at execution time. SIPE inspired this view of execution monitoring of domain-independent plans [Wilkins].

After determining the (in)validity of the propositions in x^*, the planner has gained some knowledge of what the actual state of the world must be. That is, after trying to ascertain the validity of the set of propositions x^*, we can assert that the actual situation x' must lie in some region in $\mathcal{C}$. This region is in general a function of x^*. Let $B(x^*)$ denote the region which can be inferred by determining the validity of the x^*. Since $x' \in B(x^*)$, we therefore have

$$x' \in B(x^*) \cap F_\psi(x). \qquad (*)$$

by combining the forward projection estimation of the effects of ψ, and the execution monitoring of x^*.

Now, assume that the planner has some deductive machinery. We model this mechanism as a function of the partial characterization of the world. The value of the function is a more restricted characterization. That is, the closure operation takes as its argument a region such as $(*)$, and returns a new region contained in $(*)$. We write the closure function as $\wp$, and so we have

$$x' \in \wp(B(x^*) \cap F_\psi(x)) \subseteq B(x^*) \cap F_\psi(x).$$

So formally $\wp$ is a partial self-map of the powerset of 2^Y. This notation is too general to be perspicacious. In practice, $\wp$ would also be a function of certain propositions which we wished to "emphasize," such as the effects of ψ, or the x^*. (Compare the simple example of a partial deductive mechanism we started with). This could be added to the notation, but we won't do it here.

Now, after executing just one motion with incomplete deductive closure, the actual situation x' will only be known to lie in some region R. For this reason, we must generalize forward projections to take regions as arguments. This is easy; we define

$$F_\psi(R) = \bigcup_{x \in R} F_\psi(x)$$

and note that it represents that the proposition

$$\bigvee_{x \in R} F_\psi(x)$$

is necessarily true. Now, a *goal* is obviously just a conjunction of propositions. It represents a region G in situation space $\mathcal{C}$. We are now ready to answer the question of when a planner can recognize that a goal has been achieved by a particular action:

Theorem: *A planner, having a partial deductive mechanism characterized by $\wp$, execution monitoring characterized by $B(x^*)$, and forward projection characterized* *by F_ψ, can recognize that an action ψ achieves a goal G when executed from any situation within region R, if, and only if*

$$\wp(B(x^*) \cap F_\psi(R)) \subseteq G. \qquad (**)$$

By importing the reachability and recognizability notions of [LMT,E], we have been able to provide a fairly precise characterization of partial deductive mechanisms. In particular, the theorem tells us exactly when a plan step achieves the goal in such a manner that the deductive machinery of the planner can recognize having won.

It remains to generalize (**) to n-step plans. This is done as follows. Given two plan steps ψ, θ, define their composed forward projection by

$$F_{\psi,\theta}(R) = F_\theta(F_\psi(R)).$$

For 2-step plans, (**) now becomes simply

$$\wp(B(x^*) \cap F_{\psi,\theta}(R)) \subseteq G.$$

The generalization to n-step plans now proceeds by the obvious induction.

This seems like a good start; much more work is required to make it computationally effective. This seems like a nice clean framework for investigating planners which have some deductive machinery, but which are still fast (and therefore in general incomplete). Better characterizations of the forward projections, execution monitoring, and closure functions are probably necessary before the theoretical model is practical. In this direction, in the next section we will generalize (**) to the case there there is uncertainty in the *effects* of ψ. This generalization makes execution monitoring even more important.

The model must also be extended for planning in time. Formally, I believe this can be done by simply indexing X by the model of time. It would be interesting for future research to determine whether there exist constraints in planning with time which could be exploited by algorithms for computing reachability and recognizability.

3.4. Failure Analysis

When a planner has incomplete deductive closure, it must deal with uncertainty about the state of the world. In this section, we will call this ***sensing uncertainty.*** There are other ways that uncertainty enters into planning. For example, there could be uncertainty in the effects of an action. This *control uncertainty* can also be modeled using the forward projections introduced above. With control uncertainty, the forward projection of a situation is simply all possible resulting situations (after a particular action). Planning with control uncertainty in the action effects is a much more realistic model of real-world actions. Fortunately, we do not require a great deal of machinery to incorporate it into the framework above; the recognizability theorem and (**) go through *mutatis mutandis.*

As we saw above, however, it is not always possible to guarantee plans, since the recognizability criterion given in the theorem may not hold. Since it may not be possible to find plans that are guaranteed to succeed, let us discuss ***Failure Analysis.*** Broadly speaking, failure analysis is the subfield of planning that deals with error diagnosis and recovery.

Sources

Failure analysis has been mentioned in the domain independent literature [Hayes, STRIPS McDermott, Wilkins, Z], and also in the domain of robotics [LMT, Brooks, E]. Domain-independent failure analysis has been unimpressive. Most ideas amount to dependency-directed backtracking in a graph-search domain. [Wilkins] allows the user to "type in" the value of a predicate and give it an unexpected value. A notable exception is [STRIPS]; STRIPS' "triangle tables" provided an execution monitoring capability that would admit an analysis similar to mine.

The robotics ideas are interesting, but domain-dependent. For example, Brooks' techniques appear not to generalize because they rely on the algebraic structure of the constraints. In robotics, the "situation space" C is a smooth manifold. Hence it seems possible to exploit its geometric structure for failure analysis in the case of compliant motions. I've explored this idea in [Donald]. That exploration is disjoint from and inapplicable to this discussion because of domain-specificity: The idea does not generalize to domain independent planning because it depends critically on the geometric structure of the tangent bundle to configuration space. The only overlap between [Donald] and here is a form of the failure analysis axioms, below.

The notion of "preimages" I'll discuss are due to domain-dependent work by [LMT]; preimages formalize a notion called *regression* due to [Waldinger] and later developed further by [Rosenchein, Pednault] and others. Rosenchein's formal "preimage"[21] framework is the closest thing I've found to this discussion. We do not know how to compute regressions in general; however, in the robotics domain, the geometric structure allows us to reduce their computation to difficult, but feasible computational geometry.

Domain-Independent Failure Analysis

It would be interesting to develop a domain-independent characterization of failure

[21] He has another name for it.

analysis. There is, in some sense, a "universal" solution to failure: it is (1) Make a plan, (2) Try the plan. (3) If the plan fails, then *Sense* the state of the world, *Build* a new world model, and ***Replan.*** This "solution" is fundamentally uninteresting; moreover, making it work requires at least three Ph.D. theses, addressing the general planning, sensing, and model-building problems. I'll discuss a less universal, but still interesting model of failure analysis here.

We will assume the notation and definitions from the deductive closure discussion. The basic idea is as follows: the problem is domain-independent planning with sensing and control uncertainty. The situation space is C. We are given a plan, which is guaranteed to achieve a goal recognizably when executed from some start region, say, R. However, the planner cannot guarantee that the initial situation lies in R; it may lie outside it. The idea of a failure analysis strategy is to try the plan anyway, while monitoring execution to detect failure. More precisely:

Axioms for Failure Analysis

> A failure analysis strategy should attain the goal when it is recognizably reachable, and signal failure when it is not. It should also permit serendipitous achievement of the goal. Furthermore, no plan guaranteed to terminate recognizably in the goal should ever be prematurely terminated as a failure. Finally, no plan should be terminated as a failure while there is any chance that it might achieve the goal serendipitously due to fortuitous sensing and control events.

We now introduce a domain-independent version of *preimages* [LMT] in order to explore these axioms. We'll first discuss failure analysis for linear plans, and then try to generalize to non-linear plans.

The *strong preimage* of a goal, with respect to some linear plan, is all situations in C from which the plan is guaranteed to achieve the goal recognizably (despite the sensing and control uncertainty).

The *weak preimage* of a goal, with respect to some linear plan, is all situations in C from which the plan *might possibly* achieve the goal recognizably, given fortuitous sensing and control events. [22]

Preimages will be assumed *strong* unless mentioned otherwise.

Now, we adapt the forward projection notation to linear plans. All forward projections in this section are with control uncertainty. Let $\{\theta_i\}$ be a set of plan steps that are totally ordered. If R is the start region in C, then the forward projection of R under this linear plan is defined as follows. Write the ordered plan as $\theta_1, \theta_2, \ldots, \theta_n$. The forward projection is simply

[22] More precisely, both kinds of preimages are also functions of the start region in C, since the preimage depends on knowing where you started. This is a subtle and interesting point, but there is not space to discuss it here.

$$F_{\theta_1,\theta_2,\ldots,\theta_n}(R).$$

We now derive an execution monitoring technique: Suppose that the plan consists of just one step. Denote

The start region by R,

The goal region by G.

We can immediately conclude that:

Claim: The one-step plan can be terminated as a failure if the execution monitor can recognize that the actual situation is outside G.

The recognizability criterion, of course, is (**), from the previous section. That is, if θ_1 is the plan step, we must have

$$\wp(B(x^*) \cap F_{\theta_1}(R)) \subseteq \mathcal{C} - G.$$

More generally,

Claim: An n-step plan can be terminated as a failure after the nth step if the execution monitor can recognize that the actual situation is outside G.

Can plan execution ever be terminated sooner? The answer is essentially yes.

Claim: Suppose we have a two-step linear plan, θ_1, θ_2. Denote the weak preimage of G under θ_2 by $W_{\theta_2}(G)$. The plan can be terminated as a failure after the first step, if the execution monitor can recognize that the actual situation is outside $W_{\theta_2}(G)$.

Since after θ_1, the actual situation must be within the forward projection of R (under θ_1), this means the execution monitor must be able to recognize entry into

$$F_{\theta_1}(R) - W_{\theta_2}(G).$$

An Example

We'll generalize this to a three-step linear plan $\theta_1, \theta_2, \theta_3$, and then stop. Denote the weak preimage of G under the linear plan θ_2, θ_3 by $W_{\theta_2,\theta_3}(G)$. We will abbreviate the statement "the execution monitor can recognize that the actual situation is in set S using (**)" by "$x_{actual} \in S$." Then the three step plan can be terminated as a failure under the following conditions:

(1) After the first step, if

$$x_{actual} \in F_{\theta_1}(R) - W_{\theta_2,\theta_3}(G), \tag{1}$$

(2) After the second step, if

$$x_{actual} \in F_{\theta_1,\theta_2}(R) - W_{\theta_3}(G), \tag{2}$$

or (3) after the third step, if

$$x_{actual} \in F_{\theta_1,\theta_2,\theta_3}(R) - G. \tag{3}$$

Now, it may happen that the execution monitor cannot tell whether the final situation is in G or its complement. In this case it cannot tell whether the plan has succeeded. This means the failure analysis axioms (above) are violated. On the other hand, we say that the plan $\theta_1, \theta_2, \theta_3$ *satisfies the failure analysis axioms with respect to R and G*, if the plan always* terminates because of (1), (2), (3), or

(4) The plan may be terminated as a success *after any step, if*

$$x_{actual} \in G \tag{4}$$

is recognized.

If it satisfies the failure analysis axioms, $\theta_1, \theta_2, \theta_3$ is said to be a *failure analysis plan*. To make this characterization complete, we also have to allow for serendipitous skipping of plan steps. For example, the first action could place x_{actual} in the strong preimage of the goal under θ_3. If this is recognized, then θ_3 should be executed next, skipping θ_2. (Executing θ_2, in fact, could cause failure!).

*Note: Let me emphasize that "always" means for all initial situations in R, and for all legal sensing and control events within the specified uncertainty.

Discussion

The generalization to n-step strategies proceeds straight-forwardly. This now gives us a formal procedure for failure analysis in (linear) plan execution. Computing the failure analysis criteria ((1-4) above) may be difficult; in general it is as hard as computing recognizability. It is harder still to verify whether a plan is a failure analysis plan, since this involves simulation over all possible outcomes. It is necessary to find an efficient way to do this; the only methods known so far are domain-dependent. This suggests that with our domain-independent failure analysis framework in hand, perhaps it would not be unreasonable to exploit domain-dependent constraints in order to make it effective.

It remains to generalize the failure analysis framework to non-linear plans. There is an obvious way to do this, but it is not very interesting. Briefly, it entails the following observation: given a non-linear plan, typically a linearization is chosen (perhaps arbitrarily) at execution time. ([Z] calls such a linearization a *completion*). Once a plan is linear, the failure criteria (above) can be employed in execution monitoring. If one wishes to guarantee the failure analysis plan—that is, to guarantee that the plan will satisfy the failure analysis axioms above—then one must quantify over all linearizations. This will be inherently exponential. The basic problem is that regression is primarily a linear technique. I still believe it is useful, but more research is needed to make failure analysis really viable in non-linear planning.

3.5. On Mathematical Models of Time and Space

I'll now explore mathematical models of time in space in more detail. This discussion is a continuation of the "Some Thoughts On Time" section in the review of [A]. In particular, we want to find a *continuous* model of time which satisfies Allen's Meet axioms.

I have already presented an argument for why a discrete model of time is not sufficient. Let us now pursue the "right" continuous model. Let me pause and say why I care. First of all, it is not clear, *a priori*, that there even exists a continuous model of time satisfying the Meet axioms. In attempts to construct one, a fair bit of hairy mathematics has been brought to bear: [Allen and Hayes] propose filters; [Forbus] non-standard analysis; [Fleck] constructs a family of manifolds with boundary that are "connected" by pseudo-metrics. I get worried when the obvious models (such as the real line) fail to satisfy the axioms, and the hairy mathematics doesn't seem to help. The worry is that in fact, the Meet axioms might be inconsistent with the axioms for manifolds, and there may not exist a continuous model. Of the various attempts to construct a continuous model, Fleck's is the most principled; however it is in draft and quite complicated to explain. The model I'll propose is somewhat simpler; but it may have more bugs. Simplicity has obvious benefits in trying to model naive physics.

We will proceed as follows:

(1) If time intervals are all open (or closed) intervals on the real line, then Allen's axioms cannot hold, since two such intervals cannot Meet.

One could use alternating open and closed intervals, but this seems asymmetrical. Alternatively, one could make time intervals be *half-open*, and contain their lower, but not their upper endpoint. We write such intervals as $[a, b)$. Allen points out that this introduces an undesirable asymmetry also; I'll discuss some other reasons why it is bad.[23]

In constructing the real line, one believes that the interval structure is in some sense "primitive"; it defines the topological and differential structure of the reals. Particularly in a model with "no points," the interval structure is critical in defining the properties of the real numbers. The proposal of using half-open intervals amounts to changing the interval structure. Since the usual interval structure is a basis for the topology of $\Re$, one would expect the employment of half-open intervals to induce a different topology, and perhaps a different differential structure.

In fact it does; if the half-open intervals are used as a basis for a topology on $\Re$, the resulting topology is called the *lower-limit* topology, and it is different from the usual topology on $\Re$. The resulting space is denoted $\Re_\ell$. $\Re_\ell$ has the following unpleasant properties:

[23] [Weld, Fleck] mention some of these points also.

(2) A space is called *totally disconnected* if its only connected sets are single points. $\Re_\ell$ is totally disconnected. Therefore no interval is connected. Furthermore, $\Re_\ell$ is not "continuous"—recall that we are using "continuous space" to mean *manifold.*

(3) Because of (2), all functions from $\Re_\ell$ are continuous.

(4) $\Re_\ell$ is not second-countable, so it is not a manifold. Therefore it has no obvious differential structure.[24]

Now, these are sound arguments against using the half-open intervals as a basis for a topology on $\Re_\ell$. However, we must also argue against the hybrid alternative of using the standard topology, but still making all time intervals be half-open. This argument goes as follows:

(4) Perceptual Salience. Whether or not an interval contains its endpoints is a possibly useful semantic distinction. In modelling space, whether or not a region contains its boundary is often critical. Naive physicists note that a cup (eg.) is perceived as containing its boundary, while the "air" around it is not.

(5) Compactness arguments. Allowing closed intervals is a simple way to ensure compactness, thereby ensuring that functions on these intervals will be bounded. This is useful in QP theory, where boundedness is sometimes assumed. It also allows the maximum value theorem to be employed. The half-open intervals are not compact.

(6) Consistency arguments. By allowing predicates on time intervals such as Finishes and Starts, Allen in effect allows references to the beginning and endpoints of intervals. Perhaps these are the only "points" which we should be able to refer to. The spatial analogue of "not being able to refer to endpoints" is "never being able to refer to the boundary of an object." This is clearly wrong. Now suppose we can refer to an interval's boundary. But each half-open interval contains its lower, but not its upper boundary. This seems wrong too.

It is now clear that our model of time must be continuous, and be able to support all four kinds of intervals: (a, b), $[a, b]$, $[a, b)$, and $(a, b]$.

Before we try to construct our own continuous model, let us briefly discuss some possible alternative approaches. I will focus on two: Non-Standard Analysis, and "Geometry without Points." Other techniques include the following: [Chrisman] suggests multi-resolution/multi-epsilon tolerancing techniques. Chrisman would model a cup on a table as *Meeting* if the "space" between the cup and the table was sufficiently "small." [Allen and Hayes] introduce filters; as far as I can tell, however, the effect of the filters simply to construct the integers or the real line (up to isomorphism). (This is not surprising; the real field has a number of uniqueness properties that could force this). Also see [Fleck] for a new construction to satisfy the axioms.

[24] It probably doesn't have one at all; this is harder to prove though.

An Excursion into Non-Standard Analysis

Non-standard analysis has been proposed by [Allen and Hayes, Forbus] as a possible solution to constructing a continuous model of time for the Meet axioms. [25] Let us briefly discuss the gist of this suggestion. What is non-standard analysis , and how is it applicable?

The appeal of non-standard analysis seems that one can get one's hands on the infinitesimals, and treat them as very small intervals. This is attractive, in that perhaps one could use the infinitesimals in the place of "points."

The basic trick in non-standard analysis is to construct the differential structure of the real line by an algebraic extension to the real field $\Re$. Consider adjoining a element ϵ to $\Re$ to obtain a ring $\Re[\epsilon]$ such that the relation

$$\epsilon^2 = 0 \tag{$*$}$$

is satisfied. Then we have

$$\Re[\epsilon] \cong \Re[x]/(x^2),$$

that is, $\Re[\epsilon]$ is isomorphic to the quotient ring of the real polynomials in x, modulo the ideal generated by the polynomial x^2. This is because, if ϕ is the substitution homomorphism taking a polynomial $p(x)$ to $p(\epsilon)$, then

$$\begin{tikzcd} \Re[x] \arrow[r, "\phi"] \arrow[d, "\pi"] & \Re[\epsilon] \\ \Re[x]/(x^2), \arrow[ur, "\cong"] & \end{tikzcd}$$

where π is the canonical quotient map.

This isomorphism can be used to dissect $\Re[\epsilon]$ and discover the structure of the reals with an infinitesimal adjoined. For example, note that elements of $\Re[\epsilon]$ can be written uniquely as linear expressions of the form $a + b\epsilon$; products are reduced using $(*)$. But $\Re[\epsilon]$ is not a field, since ϵ is not invertible. (Proof: suppose that α is an inverse for ϵ. Then $\alpha\epsilon = 1$, so $(\alpha\epsilon)(\alpha\epsilon) = 1 \cdot 1 = 1 = \alpha(\epsilon^2)\alpha = \alpha \cdot 0 \cdot \alpha = 0$. But $\Re[\epsilon]$ contains $\Re$, and therefore is not the trivial ring).

ϵ is supposed to play the algebraic role of an "infinitesimal" quantity. To make up for the lack of an inverse, we must adjoin the "hyper-real" numbers also; this then lets us divide by ϵ. In constructing derivatives, then, one is always careful to have infinitesimals and hyper-reals cancel out in division.

I think this ϵ behavior is essentially non-intuitive. The hyper-reals are even more so. I don't think that the infinitesimals make a good model of "very small

[25] In all fairness, [Allen and Hayes] suggest that they cannot make it work; Forbus mentions it as an aside, in a suggestion for future research.

intervals" for naive physics. Even the behavior of $(*)$ is non-intuitive. Non-standard analysis doesn't seem to yield a model for the Meet axioms, since it constructs the same differential structure for $\Re$, albeit using different axioms. Since the two models are diffeomorphic, they are homeomorphic too. How intervals in $\Re[\epsilon]$ can Meet is still unclear.

Non-standard analysis is interesting as a subfield of mathematics. However, outside of pure mathematics, it is frequently resorted to prematurely. Often the standard mathematics will in fact suffice. We will attempt to construct a continuous model which does not employ non-standard analysis.

Geometry Without Points

There is a side-current in geometry called "geometry without points" [Gerla, Gerla and Volpe, Grzegorczyk, Huntington, Nicod, Sullivan]. The idea is to build metric spaces, and thence geometry, without referring to points at all. This goal seems quite close to those of naive physics and time logics.

Perhaps in this manner an appropriate continuous model of time can be constructed. One approach involves constructing a metric space based on a set S of primitives called *solids*, a partial order $\leq$ on the solids called *inclusion*, and a real-valued *distance* function δ on $S \times S$. A collection of axioms must be satisfied, such as:

- *δ must be non-negative valued.*
- *If $U \leq W$, then for all solids V, $\delta(U,V) \geq \delta(V,W)$.*
- *$U \leq V$ iff $\delta(U',V) = 0$ for every $U' \leq U$.* [26]

A *point* in the space of solids is a minimal element of S under $\leq$. An *ϵ-point* (for positive ϵ) is a solid whose diameter is smaller than ϵ.

[A, Allen and Hayes] suggest exploiting the "granularity" of time intervals, and "treating points as very small intervals." The ϵ-points seem to serve this purpose nicely. [Gerla and Volpe] show that the non-empty regularized[27] compact subsets of a metric space (M,d) is a space of solids. In this case the distance function δ is the minimum distance (under d) between the sets, and $\leq$ is set inclusion.

The geometry without points axioms are consistent, and there exist continuous models for them. The models seem to solve the very tricky problems of "intervals as ϵ-points." I know of no other theory which expresses this idea and solves the problems it introduces consistently. I believe this is a fruitful, unexplored direction in which to pursue continuous models of time and space.

[26] There are a lot of these axioms, and a number of important consequences. I won't review them here; see the references for a simple synopsis.

[27] A set U is *regularized* if it is equal to the closure of its interior.

A Continuous Model of Time and Space

We are finally ready to develop our own model of time and space. It is continuous, satisfies the Meet axioms, and (for time) allows all four types of intervals. It is also very simple.

The primitives in the time model are the four interval types; in the space model they are the more general *manifolds with boundary*. Most workers in naive physics forget that the original intent of differential geometry was to model time and space in a natural way. I gently commend to them "The Naive Physics Manifesto Version Zero," which is my rather flip name for G. F. B. Riemann's inaugural lecture at Göttingen in 1854, *Über die Hypothesen, welche der Geometrie zu Grunde liegen* (On the Hypotheses which Lie at the Foundations of Geometry)[28] Compare the "First Naive Physics Manifesto" [Hayes]! Riemann wrote

> Thus arises the problem of seeking out the simplest data from which the metric relations of Space can be determined, a problem by which its very nature is not completely determined, for there may be several systems of simple data which suffice to determine the metric relations of Space...These data are—like all data—not logically necessary, but only of empirical certainty, they are hypotheses; one can therefore investigate their likelihood, which is certainly very great within the bounds of observation, and afterwards decide upon the legitimacy of extending them beyond the bounds of observation...
>
> ...Notions of quantity are possible only when there already exists a general concept which admits particular instances. These instances form either a continuous or a discrete manifold, depending on whether or not a continuous transition of instances can be found between any two of them; individual instances are called points in the first case and elements of the manifold in the second. Concepts whose particular instances form a discrete manifold are so numerous that some concept can always be found, at least in the more highly developed languages, under which any given collection of things can be comprehended (and consequently, in the study of discrete quantities, mathematicians could unhesitatingly proceed from the principle that the given objects are to be regarded as all of one kind). On the other hand, opportunities for creating concepts whose instances form a continuous manifold occur so seldom in everyday life that color and position of sensible objects are perhaps the only simple concepts whose instances form a multiply-extended[29] manifold.

I believe it would be premature to drop differentiable manifolds as a model of time and space. Instead, let's adapt them to the Meet axioms. The primitive objects (the interval-like things) will be manifolds with boundary in general. The concept of Meeting will rely on the measure-theoretic concept of a set "having measure zero."

[28] Translation in [Spivak].

[29] This means multi-dimensional.

We can define our continuous model using only very simple mathematics, because all the hard work—defining manifolds–has been done by Riemann.

Definitions. By the term *interval* we will mean any interval of the real line, of the form (a,b), $[a,b]$, $[a,b)$, or $(a,b]$, for $a < b$. By *manifold* we mean a locally Euclidean, second-countable Hausdorff space. A *manifold with boundary* is like a manifold in its interior. That it, its interior points have neighborhoods homeomorphic to $\Re^n$ for some fixed n. It may also (optionally) have *boundary* points, which have neighborhoods locally homeomorphic to $\Re^{n-1}$. Thus every manifold is a manifold with boundary. [30] For an arbitrary set U, denote its closure by $\overline{U}$, and its interior by U°.

I don't want to obscure the development here by defining exactly what it means for a set to have measure zero. [31] One can think of it this way (this is not completely general because I'm only giving necessary, but not sufficient conditions;[32] but it gets the idea across): in one dimension, a set has measure zero if it consists of a countable collection of points. In three-dimensions, any set of dimension less than 3 has measure zero.

Let $\mathcal{O} = \{ N_\alpha \}_{\alpha \in J}$ be a family of intervals (in the case of time) or 3-dimensional manifolds with boundary (in the case of space).[33] For A, $B \in \mathcal{O}$, define

A Primitive-Meets *B if and only if* $\overline{A} \cap \overline{B}$ *is non-empty and has measure zero.*

Allen's Meet relation for time depends crucially on the total ordering of time. Thus we must invoke the total ordering of $\Re$ to define the Meet relation for our continuous model. Let $<$ denote the total order on $\Re$. We will define a relation $\prec$ on sets. For two sets A, $B \subset \Re$, we write $A \prec B$ if, and only if, for all $a \in A$ and $b \in B$, $a < b$. $\succ$ (on sets) is defined analogously. Then

In one dimension, A Meets *B if and only if A* Primitive-Meets *B, and* $A - B \prec B - A$.

In three-dimensions, A Meets *B if and only if A* Primitive-Meets *B*.

We have the last definition because there is no total order on $\Re^n$ (inducing the usual topology) for $n \geq 2$. Therefore Primitive-Meets is defined to be be the same as Meets in 3D. Note that while in 1D Meets is anti-symmetric, in 3D it is symmetric.

Next, define

A **Contains** *B if and only if the interior of A (set theoretically) contains the interior of B.*

---

[30] These definitions are informal. More formally, the local homeomorphisms for manifolds (with or without boundary) must also have C^∞ overlap.

[31] See any elementary text, eg., [Hirsch].

[32] So, for instance, this ignores such oddities as the Cantor middle thirds set.

[33] The case for two dimensions is just like the three-dimensional case.

In one dimension, Contains is the inverse of During. In 3D, because there is no order, there can exist no equivalent of During, Starts, or Finishes. Therefore these predicates cannot be used to define a 3D containment relation. The Contains relation above will serve instead.

Now we say

A Primitive-Overlaps B *if and only if none of* $A \cap B$*,* $A - B$*, or* $B - A$ *have measure zero.*

In 1D, Overlaps can be constructed out of Primitive-Overlaps plus the order properties:

In one dimension, A Overlaps B *if and only if* A Primitive-Overlaps B*, and* $A - B \prec B - A$.

In 3D, Overlaps is defined to be equivalent to Primitive-Overlaps. Like Meet, it is anti-symmetric in 1D and symmetric in 3D.

Before and After have no analogues in 3D. In 1D, they are equivalent to $\prec$ and $\succ$ (on sets), respectively.

Starts and Finishes have no analogues in 3D. Both rely implicitly on a notion of endpoint. We capture this using the infimum and supremum.

We define

A Starts B *if and only if* $inf(A) = inf(B)$ *and* $A^\circ \subset B$.

Finish is defined analogously via the supremum.

Finally, there is the equivalence relation

$A = B$ *if and only if* A Contains B *and* B Contains A.

In 1D, under this relation, all four types of time intervals are equivalent. This means that they are equivalent from the point of view of the time logic. Obviously, to distinguish them, one would have to introduce another notion, such as congruence, which could be exploited by a qualitative reasoning system. Furthermore, there are situations in qualitative reasoning where one wishes to know that $[a, b)$ is strictly contained within $[a, b]$. This would require introducing additional relations into the time logic. At any rate, it seems straight-forward and unproblematic, now that we have a simple and consistent continuous model. Their introduction would be a useful addition to the time logic.

In 3D the Meet relationship is not fine-grained enough to be useful. For example, one also cares about the dimensionality of the intersection, and how many components it has. There is not space to develop such a representation here; it would be interesting to determine what relations were necessary for particular reasoning domains. See [Hayes] for more on this issue.

In this section we have developed a continuous model of time and space satisfying Allen's Meet axioms. It builds on the classical space/time models, and as a result we have not had to develop its topological or differential structure with our bare hands. Also, the definitions are extremely simple and involve no difficult mathematics. Finally, in 1D, there exist simple constant-time algorithms for computing whether a predicate is true of two given time intervals.

Abstract Discussion

Let me point out some aspects of the model at a level of considerable abstraction. Recall $\mathcal{O} = \{ N_\alpha \}_{\alpha \in J}$. Each N_α has the topological and differential structure of a manifold with boundary. We imagine the $\{ N_\alpha \}$ as residing in a canonical ambient space $\Re^n \times J$, for $n = 1$ or 3. To compare two elements of $\mathcal{O}$, we project them down into $\Re^n$ and examine the measure of their closures' intersection. This will define a Meet relation on manifolds with boundary in $\Re^n \times J$. This relation can be viewed as specifying certain identifications which topologize J. All computations, however, can be done in $\Re^n$.

3.6. Why Motion Planning is Harder than Domain-Independent Classical Planning

My SM thesis was on robot motion planning. I've wondered whether motion planning—which is domain-dependent—is more difficult than classical planning, which is domain independent. I'll discuss a comparison.

Chapman has pointed out that any complexity results for classical planners depend crucially on the weakness of the action representation. In this section we'll discuss another critical factor, the fact that in classical planning time is not dense.[34]

Consider the following problem: Suppose we have a robot which is a point in the plane. There are polygonal obstacles in the plane. The robot can move though free space or slide on the edges of the obstacles. The robot commands a motion by trying to move in a particular direction. A commanded motion results in a continuous trajectory. The *motion planning problem* is, given a goal region in the plane, to decide whether a particular commanded motion will ever reach the goal.

In 1984, Erdmann showed that if the number of edges is allowed to be countably infinite, then the motion planning problem is undecidable. His proof involves encoding the computation strings of a Turing machine as "channels" in the space.[35] By performing a twiddle on Erdmann's proof, I was able to show that in fact the motion planning problem was not even Turing-acceptable. Both results are found in [E].

[34] A totally ordered set $(T, <)$ is said to be *dense* if and only if for all $a, b \in T$, there exists a $c \in T$ such that $a < c < b$.

[35] The result does not assume an infinite input, however. The surfaces in the environment are input as the finite encoding of a recursive function on the plane.

I have always wondered why classical planning is recursively enumerable, but motion planning is not. Let us investigate why.

We can think of motion planning as having a conditional model of action, where the effects are conditionally dependent on the preconditions. The action is:

```
MOVE:
     Preconditions:  In(Current-Position, Goal)
     Postconditions: If True:  Halt
                        False: Continue Commanded Motion.
```

Non-standard analysis advocates can view the entire plan-verification as

```
(defun Verify-Motion-Plan (start-position motion-direction goal)
  (let ((time 0.0))
    (place-robot-at start-position)
    (do-forever
     (when (contained-in (current-position time) goal)
       (Halt))
     (command-motion motion-direction)
     (setq time (+ time *epsilon*)))))
```

where **epsilon** is the infinitesimal ϵ, *command-motion* commands a motion in *motion-direction*, and *current-position* takes the current time as an argument and returns the current position. Naturally, such iteration over an uncountable set would look ridiculous if I couldn't assure you that there exist fast computational-geometric algorithms for deciding the motion planning problem in the case of *finite* input (see [E]).

Now, when [Z] considers the conditional model of action in classical planning, he finds it makes plan *verification* $\mathcal{NP}$-hard with finite initial state. With infinite initial state and conditional actions, classical plan *generation* appears to be undecidable, but still recursively enumerable. But conditional actions and an infinite initial state (the edges) make motion planning[36] not recursively enumerable. The difference is that in motion planning the MOVE action must be executed at every point in a continuum of time. That is, if the motion starts at time 0, for all t bigger than 0, we must execute the MOVE action to see whether the goal has been achieved. In classical planning, time is modeled as the integers. Therefore time is not dense, and in any bounded interval (of $\Re$) only a finite number of actions can occur.

In other words, if the motion planning problem were merely to decide whether the goal was achieved at an *integer* time value, verification would become recursively enumerable (but still not recursive): one could simply enumerate the integer times, and check for containment in the goal at each time.

[36] i.e., both verification and generation.

The fundamental difference seems to be that *one* motion planning action occurring over an interval of time, is equivalent to an *infinity* of classical planning actions, each occurring at a point in the interval. We now formulate this as a theorem:

Theorem: *The problem of determining whether a proposition is necessarily true in a plan whose action representation is sufficiently strong to represent conditional actions, where the model of time is a linear continuum, and an action may be planned for any, and hence every instant in time, is not recursively enumerable.*

4. Conclusions

In this paper, I have tried to use [Z, AK, A] as a springboard for discussing issues in planning. These issues included

- Correctness and completeness for extended action representations,
- The relative power of planners and problems,
- A formal analysis of partial deductive closure,
- Failure analysis and execution monitoring in planning,
- Mathematical models of time and space for planning and naive physics.

I believe that these would be fruitful topics for future research; the are, in a way, a suggestion towards a planning research agenda, of What is to be Done in domain-independent planning.

Closing Thoughts: A Radical Proposal

There are many other directions to explore. For example, Tomás Lozano-Pérez has proposed an idea for classifying planning problems. He observes that in some sense the cubical blocks world with an infinite table is "trivial," in that there exists a situation in which the preconditions for any action are satisfied. This situation is obtained by placing all the blocks on the table. Let us call such a situation a *kernel*. It is not in general unique; for example, consider the case of two infinite, but distinct tables.

Similarly, the register exchange problem has a kernel, which is obtained by copying all registers given in the initial situation. The crucial fact about *this* kernel is that for any action, there exists an *inverse*. An *inverse* is an action which restores the previous state. Blocks world actions always have inverses, but this is not true in the register exchange world: information may get clobbered.

The salient characteristics of a "trivial" planning problem might be captured as follows:

A *kernel* is a set of situations such that for any situation k in the kernel, the following hold:

(1) The preconditions of any action, instantiated with any constants satisfying its codesignation constraints, are satisfied.

(2) For every action A, there exists an inverse B for it, such that k is re-achieved after executing A then B.

Then a planning domain is defined to be *trivial* if it has a kernel, and if there are fast, domain-independent algorithms for achieving the kernel from any state, and for achieving any state from the kernel. In our examples, these algorithms work by sorting based on the add- and delete-lists. Note that all these definitions are domain-independent.

This approach is attractive because it suggests we may define *subclasses* of domain-independent planning problems which are easier, but where completeness still holds. Such a classification scheme might result in a hierarchy of planning domains. I feel that developing a more fine-grained characterization of domains would be an interesting direction for future research; it could help strike a middle ground between "trivial" and "intractable."

Principal References

[A] **Allen, J.** (1984). *Towards a General Theory of Action and Time,* Artificial Intelligence, vol. 23. No. 2.

[AK] **Allen, J. and Koomen, J.** (1983). *Planning Using a Temporal World Model,* IJCAI-83.

[Z] **Chapman, D.** (1985). *Planning for Conjunctive Goals,* MIT AI Lab., AI-TR-802.

Other References

Brooks, R. A. (1982). *Symbolic Error Analysis and Robot Planning,* International Journal of Robotics Research, vol 1, no. 4, Dec., 29–68.

Chrisman, N. (1980). *Estimating Error in Categorical Maps,* Ph.D. Thesis, University of Bristol.

Donald, B. (1986). *Robot Motion Planning with Uncertainty in the Geometric Models of the Robot and the Environment,* (To appear), Proc. IEEE Intl. Conf. on Robotics and Automation, San Francisco, (April).

[E] **Erdmann, M.** (1984). *On Motion Planning with Uncertainty,* MIT AI Lab., AI-TR-810.

[STRIPS] **Fikes, R. and Nilsson, N.** (1971). *STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving,* Artificial Intelligence vol. 2.

Fleck, M. (1985). *Space with Boundaries: The Math Behind Interval MEET,* MIT AI Lab., Unpublished report.

Forbus, K. (1984). *Qualitative Process Theory,* MIT AI Lab., AI-TR 789.

Gerla, G. (1982). *La geometria senza punti,* Archimede, 3.

Gerla, G. and Volpe R. (1985). *Geometry Without Points,* American Mathematical Monthly, vol. 92. no. 10.

Grzegorczyk, A. (1960). *Axiomatizability of Geometry without Points,* Proc. Colloquium at Utrecht, Dordrecht.

Huntington, E. (1916). *A Set of Postulates for Abstract Geometry Exposed in Terms of the Simple Relation of Inclusion,* Math. Ann. lxxiii.

Kautz H. and Vilain M. (1986). *Forthcoming paper,* (to appear).

Hayes, P. (1985). *The Naive Physics Manifesto,* Ablex, in "Formal Theories of the Commensense World".

Hayes, P. (1976). *A Representation for Robot Plans,* 4th IJCAI.

Hirsch, M. (1976). *Differential Topology,* Springer-Verlag GTM 33.

[LMT] **Lozano-Pérez, T., Mason, M. T., and Taylor, R. H.** (1984). *Automatic Synthesis of Fine-Motion Strategies for Robots,* Int. J. of Robotics Research, vol 3, no. 1.

McDermott, D. (1982). *A Temporal Logic for Reasoning about Processes and Plans,* Cog. Sci. vol. 6.

Nicod, J. (1924). *La géometrie dans le monde sensible,* Paris.

Pednault, E. (1985). *Preliminary Report on a Theory of Plan Synthesis,* SRI AI Center Tech. Note 358.

Riemann, G. (1854). *Über die Hypothesen, welche der Geometrie zu Grunde liegen,* Inaugural lecture, Univ. of Göttingen.

Rosenchein, S. (1981). *Plan Synthesis: A Logical Perspective,* IJCAI-81.

Schmolze, J. (1986). *Physics for Robots,* Forthcoming Ph.D. Thesis.

Spivak, M. (1979). *A Comprehensive Introduction to Differential Geometry,* Publish or Perish, Inc..

Stefik, M. (1980). *Planning with Constraints,* Stanford Comp. Sci. Dept. Memo 80-784.

Sullivan, T. (1971). *Affine Geometry Having a Solid as a Primitive,* Notre Dame Jour. Formal Logic, XII.

Sussman, G (1973). *A Computational Model of Skill Acquisition,* American Elsevier, New York.

Tate, A. (1976). *Project Planning Using a Hierarchic Nonlinear Planner,* Dept. Artificial Intelligence Research Rept. No. 25, University of Edinburgh.

Waldinger, R. (1975). *Achieving Several Goals Simultaneously,* SRI A.I. Center Tech. Note 107.

Weld, D. (1986). *The Use of Aggregation in Causal Simulation,* Artificial Intelligence, To Appear.

Wilkins, D. E. (1984). *Domain-Independent Planning: Representation and Plan Generation,* Artificial Intelligence, vol. 22 No. 3.

Williams, B. (1984). *Qualitative Analysis of MOS Circuits,* Artificial Intelligence.

Acknowledgments

I'd like to thank David Chapman, Michael Erdmann, Margaret Fleck, David McAllester, Tomás Lozano-Pérez, and Dan Weld for valuable discussions on planning, time, qualitative reasoning, and naive physics.