Final version in H. Kyburg, R. Loui and G. Carlson (eds.), *Knowledge Representation and Defeasible Reasoning*, Kluwer Acad. Publ., Dortrecht, 1990, pp. 23-67.

Monotonic Solution of the Frame Problem in the Situation Calculus:

An Efficient Method for Worlds with Fully Specified Actions

Lenhart K. Schubert

Department of Computer Science, University of Rochester, and Department of Computing Science, University of Alberta

Abstract. The paper is concerned with the succinct axiomatization and efficient deduction of non-change, within McCarthy and Hayes' Situation Calculus. The idea behind the proposed approach is this: suppose that in a room containing a man, a robot and a cat as the only potential agents, the only action taken by the man within a certain time interval is to walk from one place to another, while the robot's only actions are to pick up a box containing the (inactive) cat and carry it from its initial place to another. We wish to prove that a certain object (such as the cat, or the doormat) did not change color. We reason that the *only* way it could have changed color is for the man or the robot to have painted or dyed it. But since these are not among the actions which actually occurred, the color of the object is unchanged. Thus we need no frame axioms to the effect that walking and carrying leave colors unchanged (which is in general false in multi-agent worlds), and no default schema that properties change only when we can prove they do (which is in general false in incompletely known worlds). Instead we use *explanation-closure* axioms specifying all primitive actions which can produce a given type of change within the setting of interest. A method similar to this has been proposed by Andrew Haas for single-agent, serial worlds. The contribution of the present paper lies in showing (1) that such methods do indeed encode non-change succinctly, (2) are independently motivated, (3) can be used to justify highly efficient methods of inferring non-change, specifically the "sleeping dog" strategy of STRIPS, and (4) can be extended to simple multiagent worlds with concurrent actions. An ultimate limitation may lie in the lack of a uniform strategy for deciding what fluents can be affected by what agents in a given domain. In this respect probabilistic methods appear promising.

1 Introduction

> "One feels that there should be some economical and principled way of succinctly saying what changes an action makes, without having to explicitly list all the things it doesn't change as well; yet there doesn't seem to be any other way to do it. *That* is the frame problem".
>
> – Pat Hayes (1987:125)

The frame problem originally surfaced within McCarthy's Situation Calculus (McCarthy 1968), when McCarthy and Hayes (1969) applied it to reasoning about goal achievement. To illustrate their approach, they considered the problem of initiating a telephone conversation. They began by writing down plausible axioms which seemed to characterize the preconditions and effects of looking up a person's telephone number and dialling that number. However, they found that they were still unable to prove that the plan "look up the number and dial it" would work, even if all the initial conditions were right (i.e., that the caller had a telephone and a telephone book, that the intended party was home, etc.). For example, the axioms provided no assurance that looking up the number would not make the caller's telephone disappear, thus voiding a precondition for dialling.

At this point, McCarthy and Hayes made a move which set the stage for all subsequent discussions of the frame problem, and proposals to solve it: they augmented their axiom for the effects of looking up a phone number, so that it asserted that the action does *not* make the caller's possessions disappear, and does *not* change the intended party's location. These, of course, are the sorts of axioms known as frame axioms.

They apparently viewed their strategy of specifying the relationships *not* changed by an action as the only one available within the Situation Calculus proper, though they deplored both its *ad hoc* character and the proliferation of axioms to which it leads:

> "If we had a number of actions to be performed in sequence we would have quite a number of conditions to write down that certain actions do not change the values of certain fluents [properties and relationships]. In fact with n actions and m fluents, we might have to write down mn such conditions."

One might add that these conditions are rather implausible in a world with multiple agents (like the one we live in). For instance, there is no assurance in real life that either the intended party, or all one's possessions will stay put while one is consulting a phone book.

Virtually all later discussions of the frame problem reiterate McCarthy and Hayes' line of reasoning, without questioning the need for frame axioms of the type suggested by them, at least within the Situation Calculus and perhaps within any classical logical framework. (See, for example, the preface and articles in (Pylyshyn 1987) and (Brown 1987).)

Yet another sort of move is available, which entirely avoids frame axioms. This is to introduce axioms about what actions are *required* to produce given types of changes. This approach was proposed for a serial world by Andrew Haas (1987). An example is the following axiom (where $holding(R, x, s)$ means that the robot is holding object x in situation s, $Result(a, s)$ is the situation resulting from carrying out action a in situation s, and $Putdown(R, x)$ is the action of R putting down x, regarded as an abstract individual; as usual, a "situation" is thought of as a possible "state of the universe"):[1]

$$(\forall x, s, s')[[holding(R, x, s) \ \wedge \ \neg holding(R, x, s') \ \wedge \ s' = Result(a, s)] \\ \rightarrow \ a \ = \ Putdown(R, x)];$$

i.e., if the robot ceases to hold an object x between situations s and s', and situation s' was obtained from situation s by act a, then a must have been the act of putting down x. (For a more versatile robot, the right-hand

[1] I will consistently use lower-case identifiers for predicates and variables, and capitalized identifiers for individual constants and functions.

side of the axiom might have allowed $a = Drop(R, x)$, and perhaps one or two other actions, as an alternative to $a = Putdown(R, x)$.) Thus, given that in a certain situation the robot holds some specific object B, and in that situation performs some action *other* than $Putdown(R, B)$, we can infer from the contrapositive that the robot still holds B after that action.[2] I will give details and argue the succinctness and other advantages of the approach in section 2.

Haas termed his axioms "domain-specific frame axioms." I will instead call axioms which specify the actions needed to produce a given type of change *explanation-closure* axioms. This reflects the fact that they supply complete sets of possible explanations for given types of change. As such (I will suggest) they are important in other areas of AI, such as story understanding. It is true that the contrapositive of an axiom like the above predicts a non-change, and in that sense resembles a "frame axiom." However, it does so on the basis of the *non-occurrence*, rather than the occurrence, of certain specific actions, and it is clear that this is not what McCarthy and Hayes, or any of the many commentators on the frame problem since then, meant by frame axioms. As I will try to show, explanation closure axioms have important advantages over (traditional) frame axioms.

In section 3, I will provide a more complete illustration of how primitive actions in a serial world can be axiomatized using explanation closure. I will include an illustration that confronts the problem of *implicit effects*. An example of an implicit effect is the change in the location of the topmost object in a stack, when the base of the stack is moved; though the effect is causally direct, its detection may require any number of inference steps. I will give examples of what can and cannot be inferred in this world, contrasting this with the more usual approaches.

Despite the emphasis in the Hayes quote on succinctness, computational

[2]Similar proposals have been made by Lansky (1987), Georgeff (1987), and Morgenstern (1988). Georgeff proposes axioms of form, "If fluent p is *not* independent of event e, then e must be one of e_1, e_2, ..., e_n". However, Georgeff's approach is non-functional and less direct than Haas', in its reliance on the notion of independence (which remains somewhat unclear). Morgenstern's *persistence rules* of form "If such-and-such actions did *not* occur at time j, then fluent p is unchanged at time $j + 1$" also depend on a non-functional view of action; further, she takes these rules as being derivable from a closed world assumption about causal rules (i.e., only changes provably caused by known actions actually occur), and that is an assumption I wish to avoid.

efficiency is of obvious importance in reasoning about change and non-change. In section 4, I will show that a default strategy which is essentially the "sleeping dog" strategy of STRIPS is deductively sound when appropriately based on explanation closure. This refutes a common assumption that monotonic solutions to the frame problem are the slowest, and that the STRIPS strategy lies somehow beyond the pale of ordinary logic.

In section 5, I will briefly explore the potential of the Situation Calculus, and the present approach to the frame problem, with respect to external events, continuous change, action composition using sequencing, conditionals and iteration, and most of all, concurrency. Note that the earlier inference about persistence of *holding* depended on the assumption that actions cannot be concurrent, so that performance of one action cannot produce changes that require other actions. Extensions to worlds with concurrent actions are possible using parallel composition of actions, along with a modified form of Haas' axioms and general axioms about the primitive parts of complex actions.

An example of a composite action is $(Costart(Walk(R,L_0,L_1),\ Walk(H,L_2,$ $L_3))$ which represents concurrent walks by R and H starting simultaneously and finishing whenever both walks are done (not necessarily simultaneously). Just as in the serial case, the *Result* function is interpreted as yielding the unique new state which results if *only* the action specified by its first argument takes place. By maintaining this functional view of actions, we preserve an important property of the original Situation Calculus (exploited by C. Green, 1969): plans are terms, and can be extracted deductively from existence proofs. On the other hand, the approach may not be systematically extensible to cases where reasoning about a given situation occurs against the backdrop of a large world knowledge base. The difficulty lies in the lack of uniform principles for identifying the relevant agents and the "boundaries" of the given situation in a way that will make a functional view of action, and explanation closure, consistent with the background knowledge.

2 Explanation closure: a simple illustration and preliminary assessment

> "A weapon has been used to crush a man's skull and it is not found at the scene of the crime. The only alternative is that it has been carried away."
>
> – Isaac Asimov, *The Naked Sun*

Let us begin by going through the earlier example, adapted from Haas (1987), in more detail. We are assuming a robot's world in which the robot can walk about, paint or dye objects, pick them up and put them down or drop them, etc. He cannot perform any of these primitive actions simultaneously.[3] The immediate consequences of actions are expressed by *effect axioms* such as

A1. $(\forall x, y, s, s')[[at(R, x, s) \wedge s' = Result(Walk(R, x, y), s)]$
$\rightarrow at(R, y, s')]$

Note that the fluent literal $at(R, x, s)$ functions as a (sufficient) precondition for the success of *Walk*.

We assume that in the initial situation S_0, the robot is at location L_0 holding an object B:

$at(R, L_0, S_0)$, $holding(R, B, S_0)$

We are interested in the situation S_1 resulting from R's walking from L_0 to L_1:

$S_1 = Result(Walk(R, L_0, L_1), S_0)$

Specifically, we wish to show that R is still holding B in S_1:

G1. $holding(R, B, S_1)$

The possible explanations for cessation of holding are that the robot put down or dropped the object:

[3] *Primitive* actions are immediately executable, requiring no further elaboration or decomposition into lower-level actions (though they may require execution monitoring to see whether, in fact, they run their course as expected). All practical planning systems seem to recognize such a level of primitive actions, even though the choice of where to "draw the line" is rather arbitrary.

A2. $(\forall a, x, s, s')[[holding(R, x, s) \wedge \neg holding(R, x, s') \ \wedge \ s' = Result(a, s)]$
$\rightarrow \ a \in \{Putdown(R, x), Drop(R, x)\}],$

where $a \in \{a_1, \cdots, a_n\}$ abbreviates $a = a_1 \vee \cdots \vee a = a_n$. To prove G1, we assume its negation

$\neg holding(R, B, S_1)$,

and use (A2) along with the initial conditions and the definition of S_1 to obtain

$Walk(R, L_0, L_1) \in \{Putdown(R, B), Drop(R, B)\}$.

But syntactically distinct primitive actions are not the same:

A3 (Inequality schemas). If α and β are distinct m-place and n-place function symbols $(m, n \geq 1)$ representing primitive actions, then
$(\forall x_1, \cdots, x_m, y_1, \cdots, y_n)\, \alpha(x_1, \cdots, x_m) \neq \beta(y_1, \cdots, y_n)$, and
$(\forall x_1, \cdots, x_m, y_1, \cdots, y_n)[\alpha(x_1, \cdots, x_m) \neq \alpha(y_1, \cdots, y_m) \vee$
$(x_1 = y_1 \wedge \cdots \wedge x_m = y_m)]$.

Appropriate instances of these schemas deny that a *Walk* is identifiable with a *Putdown* or a *Drop*, and this contradiction establishes the desired conclusion G1.

Note that the traditional approach would have used a set of frame axioms including

$$(\forall a, x, y, z, s, s')[[holding(R, x, s) \wedge s' = Result(Walk(R, x, z), s)] \rightarrow holding(R, x, s')]$$

and similar ones for every other action which does not affect *holding*, in place of (A2). Explanation closure axioms are more succinct than sets of such frame axioms because there are typically few actions that change a given fluent, but many fluents that are unaffected by a given action.[4] Besides, (as suggested earlier) frame axioms do not generalize to worlds with *concurrent* actions. For example, in a world in which a robot can *simultaneously* walk and drop an object, there is no guarantee that an object held at the beginning of a walk is still held at the end.

[4]However, as Kowalski (1979: 135) showed, sets of frame axioms specifying all fluents unaffected by a given action can be collapsed by reifying fluents and quantifying over them.

The preceding succinctness claim for explanation closure axioms is quite vague. It is unlikely that it can be made fully precise, since it amounts to a claim about the structure of "natural" theories of action for real-world domains. A "natural" theory should be intuitively understandable, extensible, and effectively usable for inference. But such desiderata are hard, if not impossible, to reduce to *syntactic* constraints.

Nevertheless, the claim can be made rather plausible, if formulated *relative* to the complexity of the axiomatization of effects. The following argument is an intuitive and empirical one, in its tacit appeal to the form which effect axioms "naturally" take (in the sorts of axiomatizations familiar to AI researchers). It assumes a primitive, serial world with "explicit effects". In the next section, I will attempt a slight generalization.

Succinctness Claim 1 (for explanation closure in a primitive, serial world with explicit effects). In a natural axiomatization of a world in terms of a set of fluents and a set of nonconcurrent primitive actions, where the axioms specifying the effects of an action explicitly state which fluents become true and which ones become false, it is possible to axiomatize non-change using explanation closure axioms whose overall complexity is of the same order as that of the effect axioms.

Argument. To see the intuition behind the claim, think of the effect axioms as conditionals of form "fluent p changes if action a_1, or $a_2, \cdots$, or a_k occurs" (this may require some slight syntactic rearrangements); e.g.,

- an object changes color if it is painted or dyed (with a new color); (note that this statement may collapse two axioms, one for the effect of painting and one for the effect of dyeing);
- an object ceases to be *on* another if the robot picks it up;
- the robot changes location if he takes a walk or pushes an object; (this might again correspond to two effect axioms); etc.

Now, roughly speaking, the addition of explanation closure axioms is just a matter of changing all the "if"s to "if and only if"s. At least this is so if each of the effect axioms states *all* fluent changes engendered by the actions.

The addition of the "only if" axioms clearly will not increase the overall complexity by more than a constant factor.

I hasten to add that this is an oversimplification. Explanation closure does, in general, differ from a strict "biconditionalization" of the effect axioms – indeed, I am about to argue that this is an advantage it has over circumscriptive or nonmonotonic approaches. Nevertheless, an explanation closure axiom in a world with explicit effects typically supplies those actions as alternative explanations of a change which *produce* that change according to the effect axioms. □

One could further argue that such *relative* succinctness assures a storage complexity well below $O(mn)$, since the complexity of the effect axioms presumably lies below this. (If it did not, McCarthy and Hayes would hardly have had grounds for complaining about the potential $o(mn)$ complexity of frame axioms!) Note also that if effect axioms do not involve unboundedly many fluents for each action, their complexity should be $O(n)$, and if a fluent is not referenced in unboundedly many effect axioms, it should be $O(m)$.[5]

Being succinct, the explanation closure approach offers a viable alternative to nonmonotonic and circumscriptive approaches. Unlike nonmonotonic approaches, it does not jeopardize effective provability. Unlike circumscription, it does not create subtle problems about what to circumscribe. As Hanks and McDermott (1987) remark, finding the "right" circumscriptive theory invariably hinges on already knowing the preferred model it should deliver. I would suggest that explanation closure axioms are a natural way to express our preferences directly, at least in simple worlds. (I argue below for their naturalness).

Another crucial advantage of the approach is that it avoids overly strong persistence inferences. This point was made briefly by Haas (1987), but

[5]It would be nice to be able to replace such tentative arguments with a hard-and-fast theoretical argument to the effect that (a) the logical structure of causation is such that for the "right" choice of formal terminology (i.e., the "right" fluents and actions), effect axioms *will not* involve more than a few fluents on average; and perhaps even that (b) there is an effective procedure allowing an agent interacting with the world to converge toward such a "right" choice of terminology. Fodor (1987) seems to demand all this and more of any genuine solution to the frame problem; however, most AI researchers take a more practical view.

deserves detailed reiteration. Suppose, for example, that we want to allow for the possibility that when the robot drops an object it might break, without insisting that this will be the outcome. A natural way to approximate this situation is to make the outcome dependent on how fragile the object is, without assuming that we *know* whether it is fragile enough to break. So the effect axiom might be:

$$\begin{aligned}\text{A4.}\ (\forall x, s, s')[[&holding(R, x, s) \wedge s' = Result(Drop((R, x), s)] \\ &\rightarrow [\neg holding(R, x, s') \wedge [fragile(x, s) \rightarrow broken(x, s')]]]\end{aligned}$$

Although we won't be able to infer breakage without knowledge of fragility, we still want to assert that if an object breaks, it was dropped. This can be straightforwardly expressed by the explanation closure axiom

$$\begin{aligned}\text{A5.}\ (\forall a, x, s, s')[[&\neg broken(x, s) \wedge broken(x, s') \wedge s' = Result(a, s)] \\ &\rightarrow a = Drop(R, x)]\end{aligned}$$

Note that here (A5) cannot be derived from the corresponding effect axiom (A4) by some systematic "biconditionalization", or any other general principle. It is essentially a domain fact. (In a more realistic world, we would allow for some additional ways of breaking, such as being struck or crushed.) So, given the particulars

$\neg broken(C, S_0)$, $holding(R, C, S_0)$ and $S_1 = Result(Drop(R, C), S_0)$,

we can infer neither $broken(C, S_1)$ nor $\neg broken(C, S_1)$, and that is as it should be.

By contrast, a circumscriptive approach that minimizes the amount of "abnormality" engendered by an action (McCarthy 1984), or its causal efficacy (Lifschitz 1987), would predict $\neg broken(C, S_1)$ and hence $\neg fragile(C, S_0)$. Similarly nonmonotonic methods (Reiter 1980) would sanction this unwarranted inference. Moreover, if we are given the particulars

$\neg broken(C, S_0)$, $broken(C, S_1)$, and $S = Result(A, S_0)$,

the explanation closure approach yields the reasonable conclusion $A = Drop(R, C)$, whereas circumscriptive and nonmonotonic approaches are silent about A (given axiom (A4) but not (A5)).

Areas of uncertainty or ignorance like that concerning breakage are hard to avoid in domain theories of practical magnitude. A familiar instance of this is the "*next-to*" problem: it is hard to provide effect axioms which will supply *all* changes in *next-to* relations (without appeal to some overly precise geometrical representation). Yet circumscriptive and nonmonotonic approaches will treat the axioms as if they supplied all such changes, and as a result sanction unwarranted persistence inferences. I will return to the *next-to* problem in the next section, which contains a more elaborate "robot's world."

Finally, I claim that encoding non-change via explanation closure axioms is principled and natural, in the sense that there are reasons independent of the frame problem for invoking them. One such reason is the observation that people can come up with small sets of plausible explanations for changes of state almost instantaneously, at least in familiar domains. For example, if the grass got wet, perhaps it rained, or the sprinkler was on, or dew formed overnight, or some snow melted – and that just about covers the most likely explanations. (Similarly consider, "How did the wall come to be blue?", "Why is the sun no longer shining?", "How did John's location get changed from the ground floor to the 17th floor of his apartment building?", "How did John learn about the earthquake in Italy while having breakfast alone in his New York apartment?", "How did John gain possession of the hamburger he is eating?", "What is causing John's nose to be runny?", etc.) Endowing machines with comparable abilities would seem to require some quite direct encoding of the connection between various phenomena and their immediate causes. Furthermore, research in natural language understanding has shown that the ability to infer actions that accomplish given state changes is extremely important, and has led to postulation of knowledge structures very similar to explanation-closure axioms. For example, Schank and Abelson (1977:75) suggest that state changes deliberately brought about by human agents are associated with sets of possible actions (in their terms, sets of "plan boxes") that achieve those state changes. They assume that if a story leads to the inference that an agent will try to accomplish a state change, the further inference is warranted that he will attempt one of the associated actions. Clearly this involves a tacit closure assumption that a deliberately caused state change is normally brought about by one of a fixed set of actions.[6]

To be sure, examples of "real-world" explanation closure are generally subtler than (A2) or (A5). They vary in level of detail (scale or "grain size") and level of abstraction (see section 5), and most importantly, are "defeasible" – the standard explanations occasionally *do* fail. However, my primary concern here is with causally insulated, predictable worlds, free of booby-trapped boxes and meteor strikes. Everything of interest that occurs will be attributable to known agents. In such a setting, (non-defeasible) explanation closure works remarkably well.

3 Explanation closure in a world with implicit effects

In case of *holding*, the cessation of this relation can be directly attributed to a *Putdown* or *Drop*. Based on such examples, the "explicit effects" assumption required direct axiomatic connections from actions to all affected fluents. This requirement is hard to enforce in nontrivial worlds. For instance, suppose that a robot is regarded as "carrying" its own integral parts, anything "riding" in or on it, and anything those "riders", in turn, are carrying (cf. the "assemblies" of Haas, 1987). This is a useful notion, because an object "carried" by another changes location with it. Now in axiomatizing actions like *Walk* or *Pickup*, we do not want to explicitly specify *all* effects on objects carried (and left behind). Rather, we want these changes to follow from axiomatic connections between *holding, in, on,* etc., and *carrying.*

The following partial theory of a world with implicit effects serves several purposes. First, it shows that the explanation closure approach to the frame problem extends readily to such worlds. (The new closure axioms are (A16-A20).) Second, it provides a nontrivial setting for illustrating inference based on explanation closure. Finally, it provides the background for further discussion of the succinctness claim and the "*next-to*" problem.

A6. An object "carries" its integral parts, its riders, anything carried by

[6] The more recent work of Kautz and Allen(1986) also involves an idea that seems closely related to explanation closure: observed, reported or inferred actions are explained in terms of a set of a set of alternative, *jointly exhaustive* higher-level actions (plans). After several observations, it is often possible to deduce a unique top-level plan.

its riders, and nothing else.[7]

$$(\forall x,y,s)[carries(x,y,s) \leftrightarrow [integral\text{-}part(y,x) \vee rider(y,x,s) \vee (\exists z)[rider(z,x,s) \wedge carries(z,y,s)]]]$$

A7. "*Carries*" is irreflexive (so that by A6 and A9, *integral-part, rider, in, on* and *holding* are also irreflexive).

$$(\forall x,y,s)[carries(x,y,s) \rightarrow \neg carries(y,x,s)]$$

A8. An object carried by another is at the same place as its carrier.

$$(\forall x,y,z,s)[[carries(x,y,s) \wedge at(x,z,s)] \rightarrow at(y,z,s)]$$

A9. An object is a rider on another iff it is in, on, or held by it.

$$(\forall x,y,s)[rider(y,x,s) \leftrightarrow [in(y,x,s) \vee on(y,x,s) \vee holding(x,y,s)]]$$

A10. "*in*" corresponds to one or more nested in_0's.

$$(\forall x,z,s)[in(x,z,s) \leftrightarrow [in_0(x,z,s) \vee (\exists y)[in(x,y,s) \wedge in_0(y,z,s)]]]$$

A11. *Paint* has the expected effect, if the robot is next to a paintbrush, paint of the right hue, and the object to be painted, and isn't holding anything.

$$(\forall x,b,c,p,s,s')[[next\text{-}to(R,x,s) \wedge next\text{-}to(R,b,s) \wedge next\text{-}to(R,p,s) \wedge brush(b) \wedge paint(p) \wedge hue(p,c) \wedge \neg(\exists y) holding(R,y,s) \wedge s' = Result(Paint(R,x,c),s)] \rightarrow color(x,c,s')]$$

A12. *Dye* has the expected effect – much like (A11).

A13. *Putdown* has the expected effect, if the robot is holding the object. (To illustrate less direct effects, effects on *in* and *on* are also included.)

[7] Nonintegral parts, such as a computer remotely controlling a robot, need not be carried by it.

$$(\forall x, y, s, s')[[holding(R, x, s) \wedge s' = Result(Putdown(R, x), x)]$$
$$\rightarrow \neg holding(R, x, s') \wedge [above(x, y, s) \rightarrow$$
$$[[container(y) \wedge smaller(x, y) \rightarrow in(x, y, s')] \wedge$$
$$[\neg container(y) \vee \neg smaller(x, y) \rightarrow on(x, y, s')]]]]$$

A14. *Pickup* has the expected effect on *holding*, if the robot is next to the object and the object is liftable.[8]

$$(\forall x, s, s')[[next\text{-}to(R, x, s) \wedge liftable(x) \wedge \neg(\exists z)holding(R, z, s) \wedge$$
$$s' = Result(Pickup(R, x), s)] \rightarrow holding(R, x, s')]$$

A15. As in the case of (A13), we might have included additional effects of *Pickup* in (A14). Alternatively, we can state additional effects separately, as in the following axiom about (successful) *Pickup*s being able to undo *carries* relations:

$$(\forall x, y, s, s')[[next\text{-}to(R, x, s) \wedge liftable(x) \wedge \neg(\exists z)holding(R, z, s) \wedge$$
$$s' = Result(Pickup(R, x), s)]$$
$$\rightarrow [[carries(y, x, s) \wedge \neg carries(y, R, s)] \rightarrow \neg carries(y, x, s')]]$$

A16. If an object ceases to be of some color y, it was painted or dyed with some color z.

$$(\forall a, x, y, s, s')[[color(x, y, s) \wedge \neg color(x, y, s') \wedge s' = Result(a, s)]$$
$$\rightarrow (\exists z)a \in \{Paint(R, x, z), Dye(R, x, z)\}]$$

A17. A change from *not holding* an object to *holding* it requires a *Pickup* action.

$$(\forall a, x, y, s, s')[[\neg holding(x, y, s) \wedge holding(x, y, s') \wedge s' = Result(a, s)]$$
$$\rightarrow a = Pickup(x, y)]$$

A18. If an object ceases to be *in* a container, then the robot must have picked up the object, or picked up something in the container carrying the object.

[8] *liftable* is here treated as independent of the agent and the given situation (e.g., whether there are "riders" on the object), but could easily be made dependent on them.

$$(\forall a, x, y, s, s')[[in(x, y, s) \wedge \neg in(x, y, s') \wedge s' = Result(a, s)]$$
$$\rightarrow [a = Pickup(R, x) \vee$$
$$(\exists z)[a = Pickup(R, z) \wedge in(z, y, s) \wedge carry(z, x, s)]]]$$

A19. If an object x comes to be *in* a container, then the robot must have put down or dropped an object z it was holding above the container, where z is smaller than the container, and either *is* x or was carrying it:

$$(\forall a, x, y, s, s')[[\neg in(x, y, s) \wedge in(x, y, s') \wedge s' = Result(a, s)]$$
$$\rightarrow (\exists z)[[z = x \vee carries(z, x, s)] \wedge holding(R, z, s) \wedge$$
$$above(z, y, s) \wedge smaller(z, y) \wedge$$
$$[a = Putdown(R, z) \vee a = Drop(R, z)]]]$$

A20. If an object ceases to be *at* a location, then the robot took a $Walk$ to some place, and either the robot is that object, or was carrying that object.

$$(\forall a, x, y, s, s')[[at(x, y, s) \wedge \neg at(x, y, s') \wedge s' = Result(a, s)]$$
$$\rightarrow (\exists z)[a = Walk(R, y, z) \wedge [R = x \vee carries(R, x, s)]]$$

This partial axiomatization lacks axioms for *on* and *next-to*, explanations for *color* or *at* becoming true, etc. While further axioms would be needed in any practical application, it is significant that even a partial axiomatization allows many reasonable conclusions about change and non-change to be drawn, as the following examples show (see also section 4). The problem of unwarranted persistence inferences, which attends circumscriptive and non-monotonic closure of incomplete theories, does not arise (at least not within settings with fully specified actions; limitations are discussed in section 5).

The following example describes initial conditions in the robot's world in which the robot is at location L_0, and is next to a blue box B containing a cup C (and perhaps other objects). In addition, there is a doormat D at location L_1, which is distinct from L_0. The problem is to show that if the robot picks up the box and walks to a location L_2, the location of the cup is changed but not the color of the box or the location of the doormat. (The descriptions "box", "cup", and "doormat" are not actually encoded in the premises, but are used for mnemonic reasons.)

Proposition 1. Given axioms (A1-A20), along with initial conditions

$at(R, L_0, S_0)$, $next\text{-}to(R, B, S_0)$, $in_0(C, B, S_0)$, $liftable(B)$, $color(B, Blue, S_0)$, $at(D, L_1, S_0)$, $L_1 \neq L_0$, $\neg(\exists z) holding(R, z, S_0)$

and plan

$S_1 = Result(Pickup(R, B), S_0)$,
$S_2 = Result(Walk(R, L_0, L_2), S_1)$

then

(a) $color(B, Blue, S_2)$, (b) $at(D, L_1, S_2)$, (c) $at(C, L_2, S_2)$.

Proof sketch.
(a): If the color of box B were not blue in situation S_1, then by (A16) the *Pickup* action which led to the situation would have had to equal a *Paint* or *Dye* action, which is impossible by (A3). Similarly we infer the persistence of B's color through the *Walk*.

(b): We assume that the doormat does *not* stay at L_1. Then by explanation closure for cessation of *at* (A20), the robot walked from L_1 to some location and either is D or carried D. But this is impossible, because the *Pickup* was no *Walk*, and the *Walk* was from L_0, which differs from L_1. (Besides, the robot is not D, and didn't carry D, because the locations of D and the robot in situation S_0 are distinct.)

(c): To prove the cup ends up at L_2, we first show that the robot ends up there, by (A1). Next, we show he ends up *holding* the box B, since the *Pickup* in the first step succeeds by (A14) and the *holding* persists through the *Walk* (by explanation axiom (A2) for cessation of holding, and the inequality schemas). Hence, we deduce by (A8) that the box ends up *at* L_2 (via the *rider* and *carries* relations, (A5) and (A6)). Next we infer by (A10) that since cup C is in_0 the box, it is *in* it, and that this relation persists through the *Pickup* and the *Walk*, using explanation axiom (A18) for cessation of *in*. (The former inference requires use of irreflexivity for *in* (A6, A7, A9), to rule out the possibility that in picking up the box, the robot lifted the cup *out* of the box along with the box!) Finally, with the box at L_2 and the cup in it, we infer by (A8) that the cup is at L_2 (via the *rider* and *carries* relations). $\Box$

So non-change, as well as change, can be straightforwardly deduced in our robot's world, without appeal to nonstandard methods. As well, it is relevant to consider what sorts of things *cannot* be inferred in this world. Suppose, for instance, we add an assumption that there is a video camera at the robot's location at the outset, i.e., $at(VC, L_0, S_0)$. We can deduce neither $at(VC, L_0, S_2)$ nor its negation, and that is as we would want. After all, the camera may or may not be attached to (or carried by) the robot.

Is the succinctness claim still tenable in such worlds with implicit effects? I submit that it is, although the evidence, even more than before, must be sought in examples (such as the one just presented) and in our intuitions about "natural" axiomatic theories.

Succinctness Claim 2 (for explanation closure in a primitive, serial world with implicit effects). In a natural axiomatization of an intuitively comprehensible dynamic world in terms of a set of situational fluents and a set of (nonconcurrent) primitive actions, it is possible to axiomatize non-change using explanation closure axioms whose overall complexity is of the same order as that of the effect axioms *plus* the axioms relating primary fluents (those explicitly connected to actions) to secondary ones.

Argument. In this case, "biconditionalizing" effect axioms of form "fluent p changes if action a_1, or $a_2, \cdots$, or a_k occurs" will provide explanation closure axioms for the primary fluents only (in approximate form). Do we also need closure axioms for secondary fluents? The preceding example suggests that secondary fluents will often have *definitions* in terms of primary ones (see *carries* and *rider* in (A6) and (A9)). Changes in such fluents are fully determined by – as well as explained by – changes in the relevant primary fluents. For example, if an object ceases to be a rider on another, we can infer from (A9) that if it was previously *in*, *on* or held by the other object, that relationship ceased; hence we can infer what action (or possible actions) must have occurred. So it appears that separate closure axioms will often be redundant for secondary fluents.

But even where such axioms turn out to be necessary or convenient, the overall complexity of closure should not exceed that of other axioms. After all, for each secondary fluent at least one axiom must already be present which introduces that fluent and relates it to others. As long as explanation closure axioms do not get arbitrarily more complicated than these relational ones, the succinctness claim remains true.

Examples suggest they will not get arbitrarily more complicated. For instance, although explanation closure axioms are theoretically redundant for the *carries* and *rider* fluents of our illustration, it is convenient to have them. For explaining how a *carries* relation comes about for the robot and an object x, and how it ceases, we might say:

$$\text{A21. } (\forall a, x, s, s')[[\neg carries(R, x, s) \wedge carries(R, x, s') \wedge s' = Result(a, s)] \\ \rightarrow (\exists y)[[y = x \vee carries(y, x)] \wedge a = Pickup(R, y)]]$$

$$\text{A22. } (\forall a, x, s, s')[[carries(R, x, s) \wedge \neg carries(R, x, s') \wedge s' = Result(a, s)] \\ \rightarrow (\exists y)[[y = x \vee carries(y, x)] \wedge \\ [a = Putdown(R, y) \vee a = Drop(R, y)]]]$$

These are no more complicated than the closure axioms suggested for primary fluents like *holding* and *in*.

Indeed, it seems unlikely that a natural set of concepts for describing an *intuitively comprehensible* domain would include fluents whose changes, even under ordinary conditions, cannot be explained (at any level) in terms of a few simple alternative causes. In other words, it seems to me that having simple explanation and prediction rules for a dynamic world is what *makes* it intuitively comprehensible. $\Box$

Finally, let us return to the *next-to* problem, whose relevance to practical robot problem solving makes it a touchstone for putative solutions to the frame problem. Essentially the problem is that neither persistence nor change of *next-to* relations can be reliably inferred for all pairs of objects. For example, suppose that our robot's world contains two adjacent windows W_1, W_2 and (for whatever reason) the robot is interested in the goal

$(\exists s) next\text{-}to(R, W_1, s) \wedge \neg next\text{-}to(R, W_2, s)$.

Suppose also that the robot has a *Go-next-to* action, which is capable of taking him next to either window. (Assume for this discussion that *Go-next-to* replaces *Walk*, though it wouldn't be hard to allow for both.) But if he walks next to W_1, will he be next to W_2? Perhaps so, if the execution routines choose a place between the windows, and perhaps not, if they choose a place next to W_1 but on the far side from W_2. In such a case we do *not* want the robot to think he can achieve the above goal by starting at a place *not* next to W_2, and going next to W_1, with the conviction that $\neg \textit{next-to}(R, W_2, S_0)$ will persist. Rather, he might decide the problem is not amenable to reliable solution, or he might know some facts which will allow him to overcome the problem (e.g., he might just happen to know that if he goes next to the left portion of a window's frame, he will be next to the window but not next to any windows or doors to its right).[9]

Similarly, it would be risky to assume (as STRIPS-style robots typically do) that when the robot walks, it ceases to be *next-to* whatever stationary objects it was *next-to* at the start. After all, it may only have travelled a short distance, or along a trajectory parallel to an object (e.g., alongside a table).

One possible way of dealing with the *next-to* problem is to rely on an exact geometrical model (e.g., one which divides up the floor space into tiles, and deduces *next-to* or $\neg$*next-to* from which tiles are occupied). For this to permit the construction of reliable plans involving *next-to*, however, we have to insist that all actions available to the robot *precisely* and *predictably* determine his location. But this is just not a tenable assumption in a realistic, reasonably complex world.

Now the challenge is this: how do we avoid unsound persistence and change inferences, such as those above, while still obtaining those that *are* sound? For instance, we *do* want to infer that the robot's *next-to* relations don't change, say when he picks up, puts down, or paints an object (under a "horizontal" interpretation of *next-to*); and we *do* want to infer that *nonmoving* objects maintain their *next-to* and $\neg$*next-to* relations.

This challenge, ostensibly a very serious one for nonmonotonic and circumscriptive approaches, is easily met by explanation closure. For instance, we can state that $\textit{next-to}(R, x, s')$ becomes true only if the robot goes *next-to* an object y (possibly x itself) which is not remote from x (where, say, *remote* means beyond four times the maximum distance for being *next-to*):

$$\text{A23.}\ (\forall a, x, s, s')[[\neg \textit{next-to}(R, x, s) \wedge \textit{next-to}(R, x, s') \wedge s' = \textit{Result}(a, s)] \\ \rightarrow (\exists y)[a = \textit{Go-next-to}(R, y) \wedge \neg \textit{remote}(x, y, s)]]$$

[9]Anyone inclined to think the robot ought to just make some default assumption, such as that he'll not be next to W_2, should imagine a situation in which W_1 has its blinds drawn but W_2 does not, and the robot is aware of a sniper across the street, bent on his destruction!

This does *not* require exhaustive knowledge of what's remote from what, but if we *do* happen to know that the object the robot went to is remote from x, we can exclude x from the set of objects the robot may now be next to. Note that the axiom also permits inference of persistence of $\neg$*next-to*(R, x, s) if the robot did something *other* than a *Go-next-to*. Similarly we can add closure axioms for *next-to*(R, x, s) becoming false, and for *next-to*(x, y, s) becoming true or false for objects x, y other than the robot. (They will be much like the *at*-closure axiom, (A20).) These will capture just the persistences that are intuitively warranted by our conception of *next-to*.

The next section describes a practical and deductively sound way in which explanation closure axioms can be translated into efficient, STRIPS-like persistence inference methods.

4 STRIPS revisited: explanation closure meets the sleeping dog

The practical problem of efficiently inferring change and non-change has been discussed by many writers on the frame problem (B. Raphael, 1971, being an early example). Ideally, we would like to match the *constant-time* inference of non-change achieved by STRIPS-like systems (Fikes & Nilsson 1971). These employ the "sleeping dog" strategy: fluents referenced by the add-lists and delete-lists of operators are updated, and the rest are assumed to remain unchanged.

The idea in the following is to emulate STRIPS within the Situation Calculus by working out certain effects of plan steps, and inferring persistence via default rules. The default rules treat the "most recent" values of fluents as still correct in the current situation. One novelty is that explanation closure axioms are used to guard against overly strong persistence inferences (by flagging certain fluents as "questionable"). The default inferences are deductively sound (and in special cases, complete) relative to a domain theory which includes the explanation closure axioms.

I will first illustrate these techniques for a specific set of fluents in a slightly curtailed version of the previous "robot's world." In this case no flagging of fluents is needed, and the rules are not only sound, but also complete for fluents of form $(\neg)holding(R, \beta, \sigma)$, relative to any "certifiable" plan – one whose steps have provably true preconditions. Further, they permit constant-time persistence inference when suitably implemented.

I will then abstract from this example, and provide a general method for using explanation closure axioms as "sentries" which watch for actions that may change a given fluent. This enables flagging fluents so as to pave the way for sound (but not in general complete) default inferences.

In order to look up the "most recent" value of a fluent one needs to have *worked out* the relevant values at each step of a plan. Consequently, any formal claims about such strategies must rely on some formalized notion of the updating process.

In the *holding* example, this is accomplished by defining an initial "world" (theory) D_0 and a succession of augmented worlds D_1, $D_2, \cdots$, where each D_i incorporates D_{i-1}, a new plan step, and some logical consequences of the step. In practice, one would expect each D_i to be derived by some "forward inference" process from D_{i-1} and the ith plan step. In the example, the forward inferences have been judiciously chosen to provide explicit preconditions for any subsequent *Pickup, Putdown,* or *Drop* step, and formulas of the right sort for making sound and complete persistence inferences.

Our domain axioms will essentially be (A2) - (A19). By leaving out the *Walk*-axiom (A1) and explanation axiom for changes in *at*, (A20), we have changed the robot from a rover to a stationary manipulator. This allows us to avoid *next-to* reasoning; in fact, we can drop the situation argument from *next-to*, so that $\textit{next-to}(R, x)$ is permanently true, or permanently false, for any object x.

As another practical measure we invoke the "unique names assumption"; i.e., all constants of our theory are drawn from a set *Names*, where these are interpreted (much as in the case of action names) as having distinct denotations. This could be expressed by axiom schema $\alpha \neq \beta$, where α, β are distinct names.

An initial world description D_0 consists of (A2)-(A19) (with *next-to* changed as discussed) along with $\neg(\exists x) holding(R, x, S_0)$, any number of additional formulas which can be consistently added, and all instances of $liftable(\beta)$ and $\textit{next-to}(R, \beta)$ entailed by the rest of D_0 for constants β occurring in D_0. A plan is a set of formulas

$$S_i = Result(\alpha, S_{i-1}) \quad , \quad i = 1, \cdots, N,$$

where each $\alpha \in \{Pickup(R,\beta), Putdown(R,\beta), Drop(R,\beta), Paint(R,\beta,\gamma), Dye(R,\beta,\gamma)\}$ for some $\beta \in$ *Names* and $S_1, \cdots, S_N$ are constants distinct from each other and from all constants occurring in D_0. The augmented descriptions relative to such a plan are given (for $1 \leq i \leq N$) by

1. for $S_i = Result(Pickup(R,\beta), S_{i-1})$, $\beta \in$ *Names*, and
   $\{next\text{-}to(R,\beta),\ liftable(\beta),\ \neg(\exists z)holding(R,z,S_{i-1})\} \subset D_{i-1}$,

   let $D_i{=}D_{i-1} \cup \{S_i = Result(Pickup(R,\beta), S_{i-1}), holding(R,\beta,S_i)\}$;

2. for $S_i = Result(Putdown(R,\beta), S_{i-1})$, $\beta \in$ *Names*, and
   $holding(R,\beta,S_{i-1}) \in D_{i-1}$,

   let $D_i{=}D_{i-1} \cup \{S_i = Result(Putdown(R,\beta), S_{i-1}), \neg holding(R,\beta,S_i),$
   $\neg(\exists z)holding(R,z,S_i)\}$;

3. same as (2), with *Drop* replacing *Putdown*;

4. for α a *Paint* or *Dye* action (whose effects can be left implicit, since only *holding* relations are to be inferred by default),

   let $D_i{=}D_{i-1} \cup \{S_i = Result(\alpha, S_{i-1})\}$.

Note that in essence, each of (1) - (3) "checks" the preconditions of the action, and adds appropriate postconditions (effects). These follow logically from D_{i-1} together with the new step. For instance in (2), $\neg holding(R,\beta,S_i)$ is added as a logical consequence of the effect axiom (A13) for *Putdown*. $\neg(\exists z)holding(R,z,S_i)$ is also a consequence, though not an obvious one: it follows from the presence of $\neg(\exists z)holding(R,z,S_0)$ in D_0 (and hence D_i) and from the explanation axiom (A17) for *holding* becoming true (an inductive proof is required). It would not ordinarily be found by forward inference, but is included to secure completeness in the "sleeping-dog" proposition to follow.

Evidently, D_i does not exist if the preconditions of some step aren't provable. However, D_i exists *whenever* the preconditions for *Pickup, Putdown,* or *Drop* actions are provable (because (4) is indifferent to the preconditions of *Paint* and *Dye* steps). I will term such plans *certifiable* (relative to D_0).

As a final preliminary we note the following way of applying explanation closure axioms to multistep plans (expressed as *Result*-equations):

Serial Plan Lemma. Given an explanation closure axiom

$$(\forall a, x_1, \cdots, x_k, s, s')[[\pi(x_1, \cdots, x_k, s) \wedge \overline{\pi}(x_1, \cdots, x_k, s') \wedge \\ s' = \mathit{Result}(a, s)] \quad \rightarrow \varphi(a)],$$

where π is a negated or unnegated predicate and $\overline{\pi}$ its complement and $\varphi(a)$ a formula containing a, and a plan

$S_i = \mathit{Result}(\alpha_i, S_{i-1}), \quad i = 1, \cdots, N,$

such that $\pi(\tau_1, \cdots, \tau_k, S_0)$ and $\overline{\pi}(\tau_1, \cdots, \tau_k, S_N)$ hold (where $\tau_1, \cdots, \tau_k$ are terms), we can conclude that for some i $(1 \leq i \leq N)$, $\varphi(\alpha_i)$.

Proof. Obviously $\pi(\tau_1, \cdots, \tau_k, S_{i-1})$ and $\overline{\pi}(\tau_1, \cdots, \tau_k, S_i)$ must hold for some i, allowing application of the closure axiom. □

Sleeping-dog proposition for *holding*. Let D_N be a theory (i.e., domain theory and certifiable plan) as defined above. Then the following default rules are sound and complete for conclusions of form $\mathit{holding}(R, \beta, S_k)$ and $\neg\mathit{holding}(R, \beta, S_k)$, where $\beta \in \mathit{Names}$ and $0 < k \leq N$:

$$\frac{\mathit{holding}(R, \beta, S_i)}{\mathit{holding}(R, \beta, S_k)}, \quad \frac{\neg\mathit{holding}(R, \beta, S_i)}{\neg\mathit{holding}(R, \beta, S_k)}, \quad \frac{\neg(\exists z)\mathit{holding}(R, z, S_i)}{\neg\mathit{holding}(R, \beta, S_k)}$$

where i is the largest integer $\leq k$ such that at least one of $\mathit{holding}(R, \beta, S_i)$, $\neg\mathit{holding}(R, \beta, S_i)$, and $\neg(\exists z)\mathit{holding}(R, z, S_i) \in D_k$.

Proof.
Soundness: We need to show that if i (as defined) exists for a given $\beta \in \mathit{Names}$, then D_k entails whichever conclusions are given by the default rules. Suppose otherwise, i.e., there are β, i satisfying the premises for which a default rule gives a conclusion whose negation follows from D_k, or neither the conclusion nor its negation follows from D_k. Consider the case where $\mathit{holding}(R, \beta, S_i) \in D_k$ and $D_k \vdash \neg\mathit{holding}(R, \beta, S_k)$. Then by (A2) (explanation for *holding* becoming false) and the Serial Plan Lemma, there was a step $S_j = \mathit{Result}(\alpha, S_{j-1})$ with $\alpha \in \{\mathit{Putdown}(R, \beta), \mathit{Drop}(R, \beta)\}$ and $i < j \leq k$. By the unique-names assumption and inequality schemas (A3), this step must appear in D_k in precisely this form, i.e., as jth step of the

plan. But then by (2) and (3), $\neg holding(R, \beta, S_j) \in D_k$, contrary to the definition of i. Next consider the case where $\neg holding(R, \beta, S_j) \in D_k$ and $D_k \vdash holding(R, \beta, S_k)$. Then a contradiction is derived just as before, using (A17) (explanation for *holding* becoming true) and (1). Third, consider the case where $\neg(\exists z) holding(R, z, S_i) \in D_k$ and $D_k \vdash holding(R, \beta, S_k)$. Then the contradiction follows just as in the previous case, except for use of the fact that $\neg(\exists z) holding(R, z, S_i) \vdash \neg holding(R, \beta, S_i)$.

Now suppose β, i are such that neither the conclusion of the applicable default rule, nor its negation, follows from D_k. Consider the case where $holding(R, \beta, S_i) \in D_k$. Since $D_k \not\vdash holding(R, \beta, S_k)$, we can consistently form $D'_k = D_k \cup \{\neg holding(R, \beta, S_k)\}$. Then in this theory we can prove that there was a step $S_j = Result(\alpha, S_{j-1})$ with $\alpha \in \{Putdown(R, \beta), Drop(R, \beta)\}$ and $i < j \leq k$, and that this step must appear explicitly in D'_k, and hence in D_k, by exactly the same line of argument as before (i.e., using the Serial Plan Lemma, unique names, and (A3)); thus we arrive at a contradiction as before. We can derive contradictions from the remaining two cases (for $\neg holding(R, \beta, S_i)$ or $\neg(\exists z) holding(R, z, S_i) \in D_k$) in an exactly analogous manner.

Completeness: Assume first that $D_k \vdash holding(R, \beta, S_k)$ for some $\beta \in Names$. We need to show that i exists as defined and $holding(R, \beta, S_i) \in D_k$ (so that default inference yields $holding(R, \beta, S_k)$).[10] By the premises of the proposition, $\neg(\exists z) holding(R, z, S_0) \in D_0$, so i certainly exists. Now suppose $holding(R, \beta, S_i) \notin D_i$. Then (by the definition of i) either $\neg holding(R, \beta, S_i) \in D_i$ or $\neg(\exists z) holding(R, z, S_i) \in D_i$. In either case, by (A17) there is a step $S_j = Result(Pickup(R, \beta), S_{j-1})$ for some j ($i < j \leq k$) and this must be explicitly in D_k by the unique-names assumption and inequality schemas. By (1), applied to D_j, this contradicts the definition of i. Second, assume that $D_k \vdash \neg holding(R, \beta, S_i)$ for some $\beta \in Names$; we show i exists as defined and either $\neg holding(R, \beta, S_i) \in D_k$ or $\neg(\exists z) holding(R, z, S_i) \in D_k$ (so that default inference yields $\neg holding(R, \beta, S_k)$). The denial of this disjunction leads to $holding(R, \beta, S_i) \in D_k$, and a contradiction follows as before. $\Box$

These default rules clearly give us a fast method of inferring non-change for *holding* (or $\neg$*holding*), when we are working out the effects of a plan

[10] Of course, if i happens to be k, the "default inference" gives nothing new.

step-by-step. In fact, we can ensure the inferences will be made in constant time (on average). We store the initial and inferred instances of literals of form $holding(R, \beta, \sigma)$, $\neg holding(R, \beta, \sigma)$, $\neg(\exists z) holding(R, z, \sigma)$, where β, σ $\in$ *Names*, in a common hash table with complex key (*holding*, β). (We include *holding* as part of the key for generality, i.e., for cases where other fluents are "tracked" as well.) Note that σ (the situation constant) is ignored in the key, so that as we progress through the plan, a list of entries will be formed for each key in chronological order. The literal needed for default inference will always be at the front of the list, allowing constant-time access.

So this provides a detailed and concrete example of efficient, STRIPS-like inference in the Situation Calculus, with the additional advantage of soundness and completeness (for a certain class of formulas) relative to the underlying domain theory. Moreover, the structure of the soundness and completeness proofs suggests that such proofs will be possible for many fluents in many applications.

Nevertheless, such default propositions are not entirely trivial to formulate (in particular, with regard to what "effect inferences" should be included in the D_i) and to prove. We would much prefer to have a *general* methodology for exploiting closure axioms for STRIPS-like default inferences.

Now it turns out that the main source of difficulty in formulating and proving sleeping-dog propositions is the goal of completeness, i.e., having the default rules cover all persistence inferences of a certain form. But it is acceptable, and ultimately necessary, to relax this constraint. It is acceptable because losing a few of the fast persistence inferences need not seriously degrade average performance. It is ultimately necessary in an unrestricted first-order theory because the forward inferences (from actions to resultant changes) needed to support subsequent default inferences may become arbitrarily hard. Clearly deducing change by forward inferencing is worthwhile only to the extent that its costs do not exceed the resultant savings in deducing non-change. It is unclear how to trade these off in general, so I will leave the issue open in the following, concentrating instead on the issue of soundness.

As soon as we consider incomplete inference of change, the risk of overly strong persistence inference arises: if at some point a change in a fluent occurred, but we failed to infer and register it, our default rules might mistakenly give us the old, outdated value as the current one. Fortunately, explanation closure axioms can be used to safeguard against such errors. Roughly the idea is to set them up as "sentries" on fluents, and "trigger" them when an action that may account for a change in those fluents occurs. Brief attempts to prove change or non-change are then made, and where both fail, the fluent is flagged as "questionable." This flagging blocks unsound default inferences. Since the flagging is essentially confined to "old" fluent literals referenced by explanation closure axioms and not subsequently updated using effect axioms, the total computational effort arguably remains modest.

In more detail, we begin with an initial world description D_0, including fluent formulas describing initial situation S_0. I will write an unspecified fluent formula for a particular situation S_i resulting from the ith step of a plan as $\varphi(S_i)$. S_i is understood to be the *only* constant situational argument occurring in $\varphi(S_i)$. $\varphi(S_k)$ is the result of uniformly substituting S_k for S_i. $\overline{\varphi}(S_i)$ is the negation of $\varphi(S_i)$ (with double negations eliminated). $?\varphi(S_i)$ is $\varphi(S_i)$ prefixed with "?", after removal of the negation, if any. We also define the *essential* fluents as some algorithmically recognizable class of fluent formulas for whose changes we have explanation closure axioms. For instance, these might be all formulas of form $(\neg)\pi(\beta_1, \cdots, \beta_k, S_i)$, where π is a primary fluent predicate (used in the axiomatization of the direct effects of actions), and $\beta_1, \cdots, \beta_k$ are constants. We now apply the following procedure. (The role of explanation closure axioms as "sentries" in step (4) is left implicit for the moment.)

Plan Tracking Procedure. We take account of the steps of a given plan $S_k = \mathit{Result}(\alpha_k, S_{k-1})$, $k = 1, \cdots, N$, expanding D_{k-1} to D_k for each k as follows. Note that for $k > 1$, D_{k-1} may contain "questioned" fluents.

1. Initialize D_k to D_{k-1}
2. Add $S_k = \mathit{Result}(\alpha_k, S_{k-1})$ to D_k
3. Apply effect axioms to this plan step in an algorithmically bounded way, adding new fluents $\varphi(S_k)$ to D_k. Some implicit effects may be deduced as well, as long as the computation is guaranteed to terminate. Preconditions of effect axioms at situation S_{k-1} may be verified in part by default rules, to be described.

4. Determine a subset V of the "visible" essential fluents. A fluent formula $\varphi(S_i)$ $(0 \leq i \leq k)$ is visible if none of $\varphi(S_j)$, $\overline{\varphi}(S_j)$, $?\varphi(S_j)$ are present for any $j > i$ (these would "conceal" $\varphi(S_i)$). V must include any visible, essential $\varphi(S_i)$ for which $\varphi(S_k)$ is not provable (i.e., for which $D_0 \cup \{S_j = Result(\alpha_j, S_{j-1}) \mid j = 1, \cdots, k\} \not\vdash \varphi(S_k)$). (Note that we must have $i < k$.) In other words, it must include the essential fluents whose persistence has not been proved, or cannot be proved. (This can be guaranteed by including *all* visible essential fluents, but this would defeat our purposes; more on this later.)

5. For each $\varphi(S_j) \in V$, initiate concurrent proof attempts for $\varphi(S_k)$ and $\overline{\varphi}(S_k)$, basing the former on relevant explanation closure axioms and the latter on relevant effect axioms. Again, conditions at situation S_{k-1} may be established with the aid of default rules. Terminate the computations by some algorithmic bound $T(\varphi(S_k), D_k)$. If the proof of $\varphi(S_k)$ succeeded, proceed to the next element of V (i.e., $\varphi(S_i)$ need *not* be concealed). If the proof $\overline{\varphi}(S_k)$ succeeded, add $\overline{\varphi}(S_k)$ to D_k. If both attempts failed, add $?\varphi(S_k)$ to D_k.

Having tracked a plan to step N, we would attempt to prove the goals of the plan, in the same manner as we prove preconditions in step (3). Of course, in a bounded proof attempt in unrestricted Situation Calculus, step (3) and the goal proof attempt may both terminate before a target formula is confirmed, even though it may be provable in principle. However, in the event of failed precondition or goal proofs we might well use some systematic way of increasing the computational effort in steps (3) and (5) (and the final goal proof). If our underlying proof procedures are complete, this will ensure that we will *eventually* prove the preconditions and goals, if indeed they are provable.

All this presupposes that the procedure as stated is deductively sound. This hinges entirely on the soundness of the default rules employed in steps (3) and (5). We now turn to these.

Default Lemma. For each $k \in \{i, \cdots, N\}$, at the end of step (5) of the

Plan Tracking Procedure, the following default rule

$$\frac{\varphi(S_i)}{\varphi(S_k)}$$

is sound for any essential fluent formula $\varphi(S_i)$ visible in D_k, i.e., $D_0 \cup \{S_j = Result(\alpha_j, S_{j-1}) \mid j = 1, \cdots, k\} \vdash \varphi(S_k)$.

Proof. By induction on k. The proposition is true for $k = 0$, since then all visible formulas are $\in D_0$. Assume it is true for all $k \leq k' - 1$ ($k' > 0$). Then the first $k' - 1$ cycles through steps (1)-(5) clearly add only logical consequences of $D_0 \cup \{S_j = Result(\alpha_j, S_{j-1}) \mid j = 1, \cdots, k' - 1\}$ to $D_{k'-1}$ (aside from questioned fluents). At the k'th cycle, the use of default rules in steps (3) and (5) to derive essential fluents $\varphi(S_{k'-1})$ is also sound by hypothesis. In step (5), by the definition of V every essential fluent $\varphi(S_i)$ such that $\varphi(S_{k'})$ is *not* deducible is concealed. Hence no such $\varphi(S_{k'})$ can be unsoundly obtained by default rules after step (4). □

Soundness is a minimal requirement if the plan tracking procedure is to provide an interesting alternative to STRIPS-like or other nonmonotonic methods. The other requirement is efficiency. How does the efficiency of the procedure compare to that of STRIPS-like methods? And does the use of the default rule provide gains over ordinary proofs based on explanation closure, like that of Proposition 1?

I don't think either of these imprecise questions can be made precise without confining oneself to some specific domain. That is an exercise we have already gone through (in the sleeping-dog proposition for *holding*), so my answers will not aspire to theoremhood. It appears that plan tracking can be roughly constant-time per plan step in STRIPS. This assumes that true preconditions can be confirmed in constant time on average (i.e., preconditions do not depend on "deeply implicit" effects), and that the fluents matched by add-list and delete-list patterns do not become arbitrarily numerous. How close does the plan tracking procedure come to this level of efficiency?

Steps (1)-(3) closely resemble precondition and effect computations for STRIPS operators, and so can reasonably be expected to be of comparably low complexity. This assumes that essential fluents correspond closely to fluents that would be referenced in STRIPS operators. It also assumes that default determination of precondition fluents will usually succeed in step (3) when it succeeds via the STRIPS (persistence) assumption; and *that* depends on steps (4) and (5), so let us turn to these.

The key question is whether in step (4), V is an *easily found, small* subset of the visible, essential fluents. If V does not become arbitrarily large (even when the number of essential fluents "tracked" becomes arbitrarily large) or arbitrarily hard to find, then step (5) will also have bounded complexity – provided that the bound T is sufficiently tight. Furthermore, if V remains small, then there will be few failures in step (3) to infer essential precondition fluents by default.

The first observation about the size of V is that it is sometimes 0. That was the point of the sleeping-dog proposition for *holding*. Essentially this was made possible by the biconditional nature of the combined effect and explanation axioms: *holding* begins iff the robot (successfully) picks something up, and ceases iff he (successfully) puts down or drops something.

But exploiting this fact required a definition of $D_1, D_2, \cdots$ tailored to the domain. How is D to be determined in general? The answer is to be sought in the explanation closure axioms. V consists of essential fluents which *may* have changed as a result of the last plan step, but have not been proved to do so. But if we have an explanation closure axiom for such a fluent, we know that the only way it *could* have changed is through the occurrence of one of the actions specified in the explanation. *This immediately rules out all the essential fluents for which the known types of explanations for change do not match the action which occurred.* This should eliminate the great majority of candidates.

Knowing that only a fraction of the visible essential fluents are candidates for V is no immediate guarantee that we can avoid sifting through them all. however, if we accept the action inequality schemas (3) and the unique-names assumption (so that "action instances that don't look the same denote distinct actions"), we can use the following sort of indexing scheme to compile V effortlessly. (i) We store (names of) explanation closure axioms in a static table with the type of fluent whose change they explain as key. (ii) We also store them in another static table with the types of actions they invoke as explanations as keys (with separate storage under each alternative explanation). (iii) Finally, we maintain a dynamic table which for each explanation closure axiom contains a list of those visible, essential fluents whose change, if it occurs, would be explained by the axiom. (These are the fluents for which the axiom serves as "sentry".) When a new essential fluent is asserted, we delete any fluent in table (iii) concealed by the new fluent (using back pointers from the fluents to the table). We look up the closure axiom relevant to the fluent in table (i), and hence store the fluent in table (iii). We can then implement step (4) of the plan tracking procedure by indexing into table (ii) for the new action α_k; we thus find the relevant "sentries" (closure axioms involving explanations which the new action instantiates), and hence retrieve the visible essential fluents potentially affected by the action from table (iii) (where we can restrict attention to those $\varphi(S_i)$ with $i < k$, as indicated in step (4)). This makes plausible the claim that STRIPS-like efficiency can be achieved, while retaining soundness.

A brief return to the *next-to* problem may help to clarify the differences between the inferences made by a STRIPS-like approach and those made by the present procedure. Let us treat fluents of form $(\neg)\mathit{next\text{-}to}(R, \beta, \alpha)$ (where β and σ are constants) as essential. We already have (A23) as possible explanation closure axiom for *next-to* becoming true, to which we might add:

$$\text{A24. } (\forall a, x, s, s')[[\mathit{next\text{-}to}(R, x, s) \wedge \neg\mathit{next\text{-}to}(R, x, s') \wedge s' = \mathit{Result}(a, s)] \\ \rightarrow (\exists y)[a = \mathit{Go\text{-}next\text{-}to}(R, y) \wedge \neg\mathit{next\text{-}to}(x, y, s)]]$$

Also, the effect axiom is

$$\text{A25. } (\forall a, x, s, s')[[\neg\mathit{next\text{-}to}(R, x, s) \wedge s' = \mathit{Result}(\mathit{Go\text{-}next\text{-}to}(R, x), s)] \\ \rightarrow \mathit{next\text{-}to}(R, x, s')]$$

We take (A3) and (A23) - (A25) as our *only* general axioms here, and assume initial situation S_0 such that

$$\neg\mathit{next\text{-}to}(R, W_1, S_0), \quad \neg\mathit{next\text{-}to}(R, W_2, S_0), \quad \mathit{remote}(\mathit{Door}, W_1, S_0), \\ \neg\mathit{next\text{-}to}(R, \mathit{Door}, S_0), \quad \mathit{next\text{-}to}(W_1, W_2, S_0)$$

Now we track the effect of "plan" $S_1 = \mathit{Result}(\mathit{Go\text{-}next\text{-}to}(R, W_1), S_0)$.
Applying effect axiom (A25):

$\mathit{next\text{-}to}(R, W_1, S_1)$

At this point, $\mathit{next\text{-}to}(R, W_1, S_1)$, $\neg\mathit{next\text{-}to}(R, W_2, S_0)$, and $\neg\mathit{next\text{-}to}(R, \mathit{Door}, S_0)$ are visible, essential fluents. ($\mathit{next\text{-}to}(W_1, W_2, S_0)$ is not essential as we have supplied no explanation closure axioms that apply; $W_1 \neq R$ by the unique-names assumption.) The first is not in V (see step (4) of procedure) since it is a current fluent ($i = k$). $\neg$*next-to*(R, W_2, S_0) leads to concurrent proof attempts for *next-to*(R, W_2, S_1) and $\neg$*next-to*(R, W_2, S_1), the former via effect axiom (A25) (which fails), and the latter via explanation axiom (A23). One way the proof strategy might proceed is by assuming *next-to*$(R, W_2, S1)$ and attempting to derive a contradiction from (A23). This yields

$$(\exists y)[\textit{Go-next-to}(R, W_1) = \textit{Go-next-to}(R, y) \wedge \neg \textit{remote}(W_2, y, S_0)].$$

By inequality schemas (A3), $W_1 = y$, so

$$\neg \textit{remote}(W_2, W_1, S_0).$$

This does not lead to contradiction; so since both proof attempts failed, the questioned fluent ?*next-to*(R, W_2, S_1) is posted.

Similarly $\neg$*next-to*$(R, \textit{Door}, S_0)$ leads to concurrent proof attempts for *next-to*$(R, \textit{Door}, S_1)$ and $\neg$*next-to*$(R, \textit{Door}, S_1)$. The former fails. The latter may again be attempted by assuming *next-to*$(R, \textit{Door}, S_1)$ and trying to derive a contradiction from (A23). This yields

$$(\exists y)[\textit{Go-next-to}(R, W_1) = \textit{Go-next-to}(R, y) \wedge \neg \textit{remote}(\textit{Door}, y, S_0)].$$

By schemas (A3), $W_1 = y$, so

$$\neg \textit{remote}(\textit{Door}, W_1, S_0),$$

contrary to a given fact. Since persistence of the robot's not being next to the door has thus been confirmed, nothing further is done: $\neg$*next-to*$(R, \textit{Door}, S_0)$ will stay visible in world description D_1 and will thus be available for default inference of $\neg$*next-to*$(R, \textit{Door}, S_1)$.

Of course, since the example only recognizes one type of essential fluent, and this is the one affected by the assumed action, it cannot serve to illustrate the claim that only a small fraction of the visible essential fluents will typically fall into subset V. What it does illustrate is the distinction the approach makes between warranted and unwarranted persistence inferences – it correctly recognizes the persistence of $\neg$*next-to*$(R, \textit{Door}, S_0)$, and correctly "questions" the persistence of $\neg$*next-to*(R, W_2, S_0). STRIPS-like, circumscriptive, and nonmonotonic approaches would fail to make this distinction.

This still leaves the question of whether the Plan Tracking Procedure, with its reliance on default inference, provides significant gains over proofs in the goal-directed style of Proposition 1.

Here the answer appears to be "not necessarily" – only in special cases.

Consider how one might try to argue the affirmative. One might, for instance, point to tasks such as letter carrying. Repetitive tasks of this type may well be of considerable interest in robotic domains. Now one might argue that a nondefault approach would have to prove after each delivery that the mail bag is still at hand, and the undelivered letters still in it. But such an argument would be erroneous. A goal-directed approach that performs inferences as needed would ignore the question of where the letters are until it was time to deliver the letter x to address y. At this point the reasoner would note that x was in the bag at the outset, that only "delivering x" can change this fact, that this action did not occur, and hence that x is still in the bag. If actions are suitably indexed (e.g., via keys like (*Deliver, Letter41*)), this inference process is a constant-time one, and hence cannot be significantly worse (in terms of order of complexity) than the default method.

Still, the default method has the advantage that in cases like the above even less work (viz., a look-up) is needed; in other words, the constant is smaller. Also, the greater explicitness of world descriptions in the default-based approach may facilitate "mental perception" processes, such as recognition of opportunities and threats. For instance, a robot planning to change a lightbulb and to hang up a calendar might "observe himself" passing close to the tool shelf in imagining his excursion to the basement to fetch a bulb. This might prompt him to obtain a hammer and nail on the same trip. "Observing" his proximity to the tool shelf requires maintenance of an up-to-date world model, one which reflects both change (his own location) and persistence (the tool shelf location). The STRIPS-inspired Plan Tracking Procedure seems well-suited to this kind of mental perception; for instance, one can imagine using "demons" which watch for opportune circumstances (relative to current goals). It would be harder to trigger such demons if the circumstances of interest could only be brought to light through persistence inference, however efficiently.

5 Possible extensions and probable limitations

Its supposed impotence *vis-à-vis* the frame problem is not the only deficiency commonly attributed to the Situation Calculus. It is also alleged to rule out concurrent actions, an independently changing world (external events), continuous change, nonprimitive and hierarchically structured actions, and other complex actions such as conditional and iterative ones.

While this range of topics is too broad for detailed consideration here, I will attempt a brief exculpation, with emphasis on the issue of concurrency. However, an interesting weakness that does emerge is that there is a kind of tension between the predicative language of propositional fluents, and the functional language of actions and *Result*. The former provides a simple means for describing change in any desired aspect of the world. The latter is *in principle* compatible with a broadly changing world, but is useful only to the extent that one adopts a localized view centered around one or a few agents. In particular, the rest of the world poses a hazard to the consistency of the functional view. So the overall picture is that the Situation Calculus is in principle much more expressive than generally assumed, but is hampered in practise by the "parochialism" of the *Result* function.

To see that the Situation Calculus does not rule out external events and agencies, think of the situations $S' = Result(A, S)$ as being the result of A *and* situation S (rather than just the result of A *in* situation S). In other words, S may be a dynamic situation, which is headed for change no matter *what* actions are initiated in it. This view allows for any sort of deterministic external change we care to describe, such as that the sun will have risen by 8:00 o'clock on any day, no matter what:

$$\begin{aligned}(\forall a, d, s, s')[[&day(d) \wedge contains(d, s) \wedge contains(d, s') \wedge \\ &\neg risen(Sun, s) \wedge\ Clock\text{-}time(s') > 8\ \wedge\ s' = Result(a, s)] \\ &\rightarrow risen(Sun, s')]\end{aligned}$$

We can even accommodate animate agencies of change, as in the arrival of buses at a bus stop. Here we might use a *Wait-for-bus* action whose "result" – thanks to the transit agency and drivers – is the presence of a bus.

However, external agencies of change do become a problem if they alter criterial fluents (those on which planned actions and goals depend) unpredictably. In such a case both effect axioms and explanation closure axioms may be invalidated. For example, if traffic on the bus route may jam, or the drivers may strike, then being at the bus stop with the fare at hand is no longer a sufficient condition for success of *Wait-for-bus*. (In other words, we encounter the *qualification problem.*) Similarly, if the money in my pocket may be arbitrarily lost or stolen, I cannot assert an axiom that its depletion requires an expenditure. Thus, I will be unable to prove the financial preconditions for boarding the bus. It would not help to include loss and theft among the possible explanations for depletion of funds, since the occurrence of these events cannot be ruled out on the grounds that some other event occurred, such as *Wait-for-bus* (or to put it differently, they weren't part of the plan).

This inability to deal effectively with a larger, more capricious world was implicit even in the earlier, sharply delimited robot's world: the closure axioms used there have highly implausible consequences if applied to the world at large. For instance, (A20), the closure axiom for cessation of *at*, together with a simple action like $S_1 = Result(Pickup(R, B), S_0)$ and the inequality schemas (A3), entails that

$\neg(\exists x, y)\ [at(x, y, S_0) \wedge \neg at(x, y, S_1)]$,

i.e., nothing moved (horizontally) between S_0 and S_1. While this is a reasonable conclusion within a restricted robot's world, it is not reasonable in a world where numerous external agencies are active concurrently with the agent of interest. One way of achieving greater realism would be to place restrictions on the variables of the closure axioms. For instance, we might say that when any one of a *certain set of objects* (nondiminutive ones within the setting of interest) ceases to be *at* a location, then the robot walked, and is that object or carried it. However, it is unclear in general how to formulate such variable restrictions in a principled, uniform manner. Even agents physically remote from an object may be able to affect it (cf. Georgeff 1987).

Despite these limitations, the fact remains that the Situation Calculus in principle admits external events.

Before moving on the the next supposed deficiency of the Situation Calculus, let us recall that it subsumes first-order logic. As such it allows the formation of complex action terms from simpler ones. This compositional potential has generally been overlooked (but see Kowalski 1986, Kowalski and Sergot 1986, and Morgenstern 1987). All of my remaining suggestions hinge on modifying or combining actions by means of functions.

In the standard "robot's world" examples (including the ones herein) change occurs in quantum jumps. However, in formalizations based on the Situation Calculus, this is not due to a limitation of the formalism (in contrast with STRIPS, for instance), but only to tradition. We can readily attain a continuous view of what goes on during an action, using a function such as $Trunc(a, t)$ for "cutting short" action a after t seconds, if it would otherwise have taken longer. The properties of truncated actions can be axiomatized using a *Time* function on situations which is real-valued and one-to-one on any set of situations constituting a "possible history of the universe" (cf., McDermott 1982, Allen 1984). *Trunc* allows us to say, for example, that at all situations s'' during $Walk(R, x, y)$ starting in situation s and ending in s', the fluent formula $moving\text{-}toward(R, y, s'')$ holds. Moreover, a slight generalization of explanation closure axioms allows us to extend persistence reasoning to ongoing actions. For example, we can modify (A2) appropriately by stating that the only primitive actions whose *initial segments* can lead to cessation of *holding* are *Putdown* and *Drop*.

Another simple use to which functions on actions can be put is to form *sequences* of actions. (McCarthy and Hayes modelled sequencing and other control regimes by inserting expressions of the Situation Calculus into Algol programs, rather than attempting composition within the Situation Calculus). In particular, we can employ a binary *Seq* function with the obvious definition

$$(\forall a, b, s)\ Result(Seq(a, b), s) = Result(b, Result(a, s)).$$

Axioms to distinguish primitive from composite actions are easily formulated, using predicates *prim* and *comp*. Another slight amendment of explanation closure axioms will then preserve their utility: in axioms like (A16) - (A24), we include the qualification $prim(a)$ in the antecedent.

Now what makes sequences of actions interesting is the possibility of using them as "macros" (larger-scale actions) in plan reasoning. For this to be profitable, however, both effect axioms and explanation closure axioms need to be formulated at the level of composite actions. Both turn out to be possible, at least within limits. For effect axioms, we can use "lemmas" about their net effect based on effects of constituent primitive actions. For explanation closure, where there are just two levels (*prim* and *comp*) of stratification, we can use entirely separate closure axioms at the *comp* level, with actions qualified as $comp(a)$ in the antecedent. For instance, suppose we have defined *Move-object* as a 3-step macro (involving *Pickup, Walk,* and *Putdown*), along with "stationary" macros like *Empty-into, Open-blind, Unlock,* and so on. Then we can state that if an object changes location *via a comp action*, the action must be a *Move-object* (and the relocated object must be the argument, or carried by it, or is the robot, or something the robot was already carrying at the start).

Generalizing beyond two levels of stratification is certainly desirable but at this point an open problem. We need to move from the two predicates *prim* and *comp* toward a general taxonomy of actions, allowing for both composition (constructing larger-scale actions out of smaller-scale ones) and abstraction (classifying a given set of actions as being of the same abstract type). As an example of abstraction, running, walking, crawling and hopping (by humans) can all be classified as types of unmechanized travel, where the latter is in turn subsumed under (mechanized and unmechanized) travel. Preliminary research suggests that persistence reasoning based on explanation closure axioms carries over to this setting, with the requirement that "action inequality reasoning" based on schemas (A3) be replaced by "action exclusion reasoning" (e.g., the incompatibility of running and walking).

One possible weakness of the Situation Calculus that emerges from a consideration of action abstraction is its somewhat counterintuitive distinction between "deterministic actions" – those (reified) actions which lead to a unique successor state via *Result* – and abstract actions – those describable only by *predicates* over (reified) actions. This flaw apparently cannot be remedied without substantial reformulation of the calculus (e.g., in terms of a result-*relation* over actions and pairs of situations) or without losing the advantage of having plans expressed as terms, allowing their deductive extraction in the manner of Green.

Conditional actions and iteration can also be introduced with the aid of composition functions such as $If\ (test, action)$ and $While(test, action)$. The details would take us too far afield, but three things are worth pointing out. First, preconditions for conditional actions must take account of the agent's knowledge about the truth of *test*, to avoid an assumption of omniscience and the risk of paradox; consider, for instance, *If*(*Goldbach-conjecture*, *Say-yes*(R)) or *If*(*Committed-to-saying-yes*(R), *Say-no*(R)) (cf., Manna and Waldinger 1987, Morgenstern 1987). Second, tests are reified propositions about situations and as such appear to call for duplicating the entire logic within its functional notation, including quantifiers and connectives (e.g., consider "test whether there is a blue cup in every box"). This is feasible (McCarthy 1979), but to my mind not very attractive. The third point is that at least if we limit ourselves to "tests" which fit into our taxonomy of actions (e.g., *prim* and *comp* in simple cases), explanation closure can be used to prove persistence through conditionals and loops – though naturally both change and nonchange inference can become quite complicated in proofs by cases or by induction.

Finally, I will consider concurrency at somewhat greater length. As before, the key is action composition, in this case by parallel combinators. I will restrict myself to one for the moment, $Costart(a_1, a_2)$ which is the action consisting of simultaneously started actions a_1, a_2, and which terminates as soon as both are done (not necessarily at the same time). a_1 and a_2 need not be independent of each other, i.e., the effect of each may depend upon the co-occurrence of the other (as, for example, in cooperative lifting and carrying of a sofa). However, I will not concern myself with reasoning about interdependent actions here.

It is important to understand the intuitive interpretation of the expression

$Result(Costart(a_1, a_2), s)$.

Just as in the case of $Result(a, s)$, this is the resultant situation when *only* the action specified by the first argument (in this case, $Costart(a_1, a_2)$) takes place. This incidently does *not* preclude external change any more than in the serial case. The notation simply says that the concurrent actions a_1 and a_2 are the only ones carried out by the agents of interest – those who from our chosen perspective generate the space of possible future histories (while any other sources of change can only be accommodated predictively).

The following example will serve to illustrate reasoning about persistence (and change) in a world with concurrent actions. In a room containing a man, a robot and a cat as the only potential agents, the only actions are that the man walks from one place to another, while the robot picks up a box containing the (inactive) cat and walks to another place. So the initial conditions (in part) and the plan are as follows:

$$at(R, L_1, S_0),\ next\text{-}to(R, B, S_0),\ in_0(C, B, S_0),\ liftable(B),$$
$$color(C, Ginger, S_0),\ at(H, L_0, S_0),\ \neg(\exists z) holding(R, z, S_0)$$

$$Hplan = Walk(H, L_0, L_3)$$
$$Rplan = Seq(Pickup(R, B),\ Walk(R, L_1, L_2))$$
$$Plan = Costart(Hplan, Rplan)$$
$$S_3 = Result(Plan, S_0)$$

Our goal is to show that the cat retains its ginger color:

(a) $color(C, Ginger, S_3)$

Since we will need to reason about the primitive parts of composite actions, we will use the following postulates.

A25. *Walk, Pickup,* etc. are primitive: for α an n-place function $\in \{$*Walk, Pickup, Paint, ...*$\}$,

$$(\forall x_1, \cdots, x_n)\ prim(\alpha(x_1, \cdots, x_n))$$

A26. A primitive part of two concurrent actions is a primitive part of one or the other.

$$(\forall x, y, z)[prim\text{-}part(x,\ Costart(y, z)) \\ \rightarrow [prim\text{-}part(x, y) \vee prim\text{-}part(x, z)]]$$

A27. Similarly for sequences of actions

$$(\forall x, y, z)[prim\text{-}part(x,\ Seq(y, z)) \\ \rightarrow [prim\text{-}part(x, y) \vee prim\text{-}part(x, z)]]$$

A28. A primitive part of a primitive action is identical with it.

$$(\forall x, y)[[prim\text{-}part(x, y) \wedge prim(y)] \rightarrow\ x = y]$$

To prove color persistence, we will use the following variant of closure axiom (A16):

A29. If an object ceases to be of color v in the course of a plan, that plan contains a primitive part which is the action of painting or dyeing the object some color w.

$$(\forall p, y, v, s, s')[[color(y, v, s) \wedge \neg color(y, v, s') \wedge s' = Result(p, s)] \\ \rightarrow (\exists x, a, w)[a \in \{Paint(x, y, w), Dye(x, y, w)\} \wedge \\ prim\text{-}part(a, p)]]$$

We can now prove our goal (a) by assuming it is false and applying (A29) with s' and p instantiated to S_3 and $Plan$ respectively. We infer that for some agent x, x painted or dyed the cat and this action is a primitive part of $Plan$. Then this action is also a primitive part of *Hplan* or *Rplan* by (A26). Hence it is a primitive part of $Walk(H, L_0, L_3)$, $Pickup(R, B)$, or $Walk(R, L_1, L_2)$ by (A27). By (A25) and (A28) the painting or dyeing action is identical with one of these three actions, contrary to the inequality schemas.

This proof (and its axiomatic basis) is very simple, and that is the primary point of the illustration. However, we would also like to confirm that *change* can be inferred in such a setting, based on reasonable success criteria for the concurrent actions involved. As in the case of serial worlds, this is a little harder than inferring persistence.

For actions which have their usual preconditions satisfied, I will take spatiotemporal disjointness of their "projected paths" as a sufficient condition for their successful concurrent execution.[11] $Path(a, s)$ can be thought of as a time-varying spatial region, namely the region which the agent of action a and all the objects it "carries" is expected to occupy from $Time(s)$ onward, if a is the *only* action initiated in situation s or beyond. Projected paths are assumed to be adhered to in the performance of an action as long as any actions concurrent with it are independent of it.

To state these assumptions formally, we need to think of situations (and time) as changing continuously throughout actions, and to provide a way of referring to portions of plans preceding or following some intermediate situation at which a component action ends. For the preceding portion, we define $Costart_1(p, q)$ as the action which consists of running p to completion while running q concurrently, cutting it off if it has not yet finished when p is done. (As in the case of $Trunc$, this does not necessarily entail an actual cutoff, but just that *Result* applied to this action will return the situation at the point where p finishes.) We will later define $Remainder(p, q, s)$ as the "left-over" portion of p.[12]

Let us prove that the cat ends up in the same final location as the robot; i.e.,

(b) $at(C, L_2, S_3)$

introducing further axioms as needed. We begin by showing that R's *Pickup* succeeds. The modified effect axiom for *Pickup* is

$$\text{A30. } (\forall a, x, y, p, s, s')[[\textit{next-to}(x, y, s) \wedge liftable(y) \wedge \neg(\exists z) holding(x, z, s) \\ \wedge\ a = Pickup(x, y)\ \wedge\ compatible(a, p)\ \wedge$$

[11]Spatiotemporal disjointness is a special case of disjoint "resource" use, if one conceives of resources broadly as including occupiable regions of space. Disjoint resource use is often a sufficient condition for compatibility of concurrent actions, though not a necessary one.

[12]In the same vein, one can delay, vacuously extend, and truncate actions, using a vacuous action $Passtime(t)$ in $Seq(Passtime(t),\ p)$, $Costart(Passtime(t),\ p)$, and $Costart_1(Passtime(t),\ p)$.

$$s' = \mathit{Result}(\mathit{Costart}_1(a,p),s)]$$
$$\rightarrow\ \mathit{holding}(x,y,s')]$$

This illustrates the generalization of effect axioms to worlds with concurrent actions. Note that the result of the action is considered in the context of an arbitrary concurrent plan p.

To apply this axiom to the robot's *Pickup* action in the context of the man's *Walk,* we need to establish the compatibility of the two actions. To minimize geometrical complexities, let us assume that we are able to calculate "action corridors" for $\mathit{Pickup}(R,B)$ and $\mathit{Walk}(H,L_1,L_2)$ independently of the situation in which they are attempted, except for being given the location of R in the *Pickup* (i.e., L_1). This is plausible if corridors are "generously" defined so as to allow for "elbow room" and as large a collection of objects as R or H are capable of carrying. (In practice the corridors might be generalized cylinders based on the geometry of the room and the agents, plus clearance.) By definition the projected path of any *Pickup* feasible in isolation will be confined to the *Corridor* for that *Pickup*, and similarly for the projected path of a *Walk*:

A31. $(\forall a,u,x,y,s)[[\mathit{at}(x,u,s)\ \wedge\ \mathit{next\text{-}to}(x,y,s)\ \wedge\ \mathit{liftable}(y)\ \wedge$
$\neg(\exists z)\mathit{holding}(x,z,s)\ \wedge\ a = \mathit{Pickup}(x,y)]$
$\rightarrow\ \mathit{confined\text{-}to}(\mathit{Path}(a,s),\ \mathit{Corridor}(a,u))]$
$(\forall a,x,y,z,s)[[\mathit{at}(x,y,s)\ \wedge\ a = \mathit{Walk}(x,y,z)]$
$\rightarrow\ \mathit{confined\text{-}to}(\mathit{Path}(a,s),\ \mathit{Corridor}(a,y)]$

Call the relevant action corridors *Corridor-R-Pickup* and *Corridor-H-Walk*, and assume they are disjoint regions of space:

A32. $\mathit{Corridor}(\mathit{Pickup}(R,B),\ L_0) = \mathit{Corridor\text{-}R\text{-}Pickup}$
$\mathit{Corridor}(\mathit{Walk}(H,L_0,L_3),\ L_0) = \mathit{Corridor\text{-}H\text{-}Walk}$
$\mathit{disjoint}(\mathit{Corridor\text{-}R\text{-}Pickup},\ \mathit{Corridor\text{-}H\text{-}Walk})$

Clearly the antecedents in (A31) are satisfied by $a = \mathit{Pickup}(R,B)$ and $a = \mathit{Walk}(H,L_0,L_3)$ respectively, and so we can conclude with the aid of (A32) that their projected paths are confined to the above-mentioned corridors. This finally puts us in a position to infer their compatibility, using

A33. $(\forall a_1,a_2,c_1,c_2,s)[[\mathit{confined\text{-}to}(\mathit{Path}(a_1,s),c_1)\ \wedge$

$$confined\text{-}to(Path(a_2, s), c_2) \ \wedge \ disjoint(c_1, c_2)]$$
$$\rightarrow \ compatible(a_1, a_2, s)]$$

The conclusion is $compatible(Pickup(R, B),\ Walk(H, L_0, L_3),\ S_0)$, and so we can instantiate (A30) and conclude that the *Pickup* succeeds, i.e., $holding($ $R, B, S_1)$, where $S_1 = Result(Costart_1(Pickup(R, B),\ Walk(H, L_0, L_3)),\ S_0)$.

To show that the robot's *Walk*, initiated right after the *Pickup*, succeeds, we begin by defining $Remainder(p, q, s)$ as a function which returns the part of p "left over" if $Costart_1(q, p)$ is executed in situation s; i.e.,

A34. $(\forall p, q, s)\ Result(Costart(p, q),\ s) =$
$Result(Seq(Costart_1(q, p),\ Remainder(p, q, s)),\ s)$

(A suitable null element can be used when nothing is left over.) The reason for having a situation argument in the *Remainder* function is that the part of p left over when q finishes in general depends on initial conditions. In addition, a $Tail$ function will serve to return the remainder of a path, starting at a specified time. Then a required axiom about conformity between actual and projected paths, in the case of compatible concurrent actions, can be stated as follows:

A35. $(\forall p, q, r, s_0, s)[[compatible(p, q, s_0) \wedge r = Remainder(p, q, s_0) \wedge$
$s = Result(Costart_1(q, p), s_0)]$
$\rightarrow \ [Path(r, s) = Tail(Path(p, s),\ Time(s))]]$

This says that if a plan p has been partially executed concurrently with another *compatible* plan till the latter was done, then the projected path for the remainder of p is unchanged from the original projection (apart from the absence of the initial path segment already completed). Thus we can use the previously inferred compatibility of R's *Pickup* and H's *Walk* to calculate the projected remainder of H's *Walk*, namely,

$Tail(Path(Walk(H, L_0, L_3),\ S_0),\ Time(S_1))$.

We assume that a situation reached from another via an action is temporally later, so this "tail" path will be a part of the complete *Walk*-path. Since the latter is confined to *Corridor-H-Walk*, it is clear (without going into further detail) that the former is also. So, assuming

A36. $Corridor(Walk(R, L_1, L_2),\ L_1) = Corridor\text{-}R\text{-}Walk$
$disjoint(Corridor\text{-}R\text{-}Walk,\ Corridor\text{-}H\text{-}Walk)$,

we can confirm the preconditions for R's *Walk*, in

$$\text{A37. } (\forall a, x, y, z, p, s, s')[[at(x, y, s) \ \wedge \ a = Walk(x, y, z) \ \wedge \\ compatible(a, p) \ \wedge \ s' = Result(Costart_1(a, p), s)] \\ \rightarrow \ at(x, z, s')]$$

At least, we will be able to confirm those preconditions if we can derive the persistence of the robot's location during the *Pickup*, i.e., $at(R, L_1, S_1)$. But this follows easily from a closure axiom for change in *at* similar to (A29) and the primitive-part axioms (A26) and (A28).

It then remains to track the location of the cat as it gets picked up and moved along with the box. This need not detain us, since it is completely analogous to the proof of Proposition 1(c). (Of course, all additional effect axioms and explanation axioms need to allow for concurrent plans in the manner of (A29), (A30) and (A37). Also, some axioms are needed for relating alternative ways of decomposing composite plans in terms of *Costart,* $Costart_1$*, Seq,* and *Remainder.*) □

Clearly, the main complication in tracking change has been the establishment of compatibility between concurrent actions. This was done by the rather crude device of assuming that action paths are confined to disjoint "corridors". Even that was a little tedious, suggesting (unsurprisingly) that the Situation Calculus is not well-suited to reasoning about detailed geometrical and kinematic relationships – at least not without supplementation by specialized data structures and algorithms.

My main objective, however, has been to demonstrate the ease of proving non-change, using explanation closure in a world with concurrent actions. Generalization of STRIPS-like plan tracking methods to worlds with concurrent actions remains an open problem. However, I see no serious obstacle to doing so at least in cases where the chronological ordering of the start and end points of the set of concurrent actions can be inferred, and concurrent actions are independent of each other.

Finally, a few words are in order on McCarthy and Hayes' telephone problem, with which I started. In a sense, this is simpler than my robot-and-the-cat problem, since it involves no concurrency (look up the number and then dial it) and hence requires no action-compatibility reasoning. If we are prepared to posit such "primitive" actions as *Lookup-number* (x, y), *Dial-number* (x, y), *Carry-off* (x, y), and *Leave-home* (x), providing effect and closure axioms in terms of these actions for fluents like *know-number* (x, y, s), *has* (x, y, s), *at-home* (x, s), and *in-conversation* (x, y, s), we will have no trouble with the problem.

However, the same caveats apply as in the discussion of external agencies of change at the beginning of this section. If we are not careful about the way we qualify success conditions for actions, or variable restrictions in explanation closure axioms, our axioms will be patently false in the world at large. This is certainly something to be avoided in a general "commonsense reasoner," yet we do not at this point have a general, principled method of doing so.

I believe that the most promising research avenue in dealing with this difficulty lies in the application of probabilistic methods such as those of Pearl(1988), Bacchus(1988), Kyburg(1988), Dean & Kanazawa(1988), and Weber(1989). These methods allow one to give expression to the "statistical" aspect of our experience and knowledge of the world. For instance, people know that a penny left on the sidewalk is much more likely to stay put for a day than a dollar bill, that a car parked at night on a residential street will stay in place much longer on average than one parked on a weekday at a supermarket, and so on. In part, this knowledge is due to direct or linguistically transmitted observation, and in part it derives from related knowledge about *why*, and *how often*, people or other agents do the things which account for change. The dollar bill illustrates both aspects: we have a pretty good idea from direct observation about the density of pedestrian traffic on various kinds of streets at various times, and we also know that few people would fail to notice a dollar bill on the sidewalk, and having noticed it, fail to retrieve it. As well, we know about winds and their effects. Such "statistical" knowledge is absolutely indispensable in coping with a complex and more or less capricious world. It may even constitute the bulk of our general knowledge.

The role of this knowledge with respect to the frame problem is that it provides a stable, yet pliable base on which we can superimpose our episodic knowledge. Since this base merely supplies statistical priors, it yields to the pressure of event reports that run against the odds, replacing probable persistence with known change. Effect axioms and explanation closure axioms would be recast probabilistically in such a representation, and supplemented with direct empirical probabilities for various kinds of change (or conversely, persistence). If we regard the success of an action as a mere likelihood, given that the major preconditions are met, we avoid a futile quest for perfectly reliable preconditions. If we regard certain actions capable of effecting change as merely improbable, rather than as assuredly absent, we avoid unfounded beliefs about the lack of change in the world at large, and about the inevitable success of our plans.

Of course, the nonmonotonic theorists can reasonably claim to be striving toward just this kind of resilient, yet amendable knowledge base. There is, however, a fundamental difference between probabilistic and nonmonotonic methods of inferring persistence. According to the former, McCarthy and Hayes' phone stays put, in the absence of information to the contrary, because we know perfectly well that phones very rarely get moved (and indeed, we know *why* they don't). According to the latter, it stays put in the absence of information to the contrary simply because there *is* no information to the contrary. The former is sensitive to the statistical facts of the world (such as that the phone is much less likely to depart than the intended party at the other end), while the latter is turned entirely inwards.

6 Conclusions

I have provided evidence that explanation closure axioms provide a succinct encoding of nonchange in serial worlds with fully specified actions, and a basis for STRIPS-like, but monotonic inference of change and nonchange in such worlds. As such, they are certainly preferable to frame axioms; they also offer advantages over circumscriptive and nonmonotonic approaches, in that they relate nonchange to intuitively transparent explanations for change, retain an effective proof theory, and avoid unwarranted persistence inferences.

Furthermore, unlike frame axioms, explanation closure axioms generalize to worlds with concurrent actions. I led up to an illustration of this claim by enumerating some generally unknown capabilities of the Situation Calculus with respect to external events, continuous change, and composite actions, all of which seem compatible with explanation closure. Throughout, I adhered to the original *Result*-formalism, so as to retain the treatment of plans as terms, and hence the possibility of extracting plans from proofs.

Limitations of the Situation Calculus I noted along the way were the tediousness of reasoning about simple spatiotemporal relationships (without special methods), an unequal treatment of primitive (concrete) and abstract actions, and most importantly, the parochial view of the world enforced by the *Result*-formalism. It works well only for domains in which the actions capable of effecting salient change are fully and reliably known. I suggested that probabilistic methods offer the best hope of overcoming this limitation.

Directions for further research are generalizations of the results (especially the "sleeping dog" strategy) to more complex theories of the world (with external events, continuous change, higher-level actions, and concurrency), investigation of planning (as opposed to mere "plan tracking") using deductive or other methods, and the study of all of these issues within a probabilistic framework.

Acknowledgements

I am grateful to Scott Goodwin and Randy Goebel for providing astute criticisms and important pointers to the literature when this work was in the early stages. Others who provided valuable comments and suggestions were James Allen and several members of a graduate class at the University of Rochester, especially Jay Weber and Hans Koomen. The paper would have languished in semicompleted state without the generous and timely help of Chung Hee Hwang on many aspects of the paper, both small and large. The initial research was supported by the Natural Sciences and Engineering Research Council of Canada under Operating Grant A8818.

References

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

[Bacchus, 1988] F. Bacchus, "Statistically founded degrees of belief," In *Proc. of the 7th Bienn. Conf. of the Can. Soc. for Computational Stud. of Intelligence (CSCSI '88)*, pages 59–66, Edmonton, Alberta, June 6-10, 1988.

[Brown, 1987] F. M. Brown, editor, *The Frame Problem in Artificial Intelligence. Proc. of the 1987 Workshop*, Lawrence, KS, Apr. 12-15, 1987. Morgan Kaufmann Publishers, Los Altos, CA.

[Dean and Kanazawa, 1988] J. Dean and K. Kanazawa, "Probabilistic Causal Reasoning," In *Proc. of the 7th Bienn. Conf. of the Can. Soc. for Computational Stud. of Intelligence (CSCSI '88)*, pages 125–132, Edmonton, Alberta, June 6-10, 1988.

[Fikes and Nilsson, 1971] R.E. Fikes and N.J. Nilsson, "STRIPS: A new approach to the application of theorem-proving to problem-solving," In *Proc. of the 2nd Int. Joint Conf. on AI (IJCAI '71)*, pages 608–620, 1971.

[Fodor, 1987] J.A. Fodor, "Modules, frames, fridgeons, sleeping dogs, and the music of the spheres," In *Z.W. Pylyshyn (1987)*, pages 139–149. 1987.

[Georgeff, 1987] M. P. Georgeff, "Actions, processes, causality," In *M.P. Georgeff and A.L. Lansky (1987)*, pages 99–122, 1987.

[Georgeff and Lansky, 1987] M.P. Georgeff and A.L. Lansky, editors, *Reasoning about Actions and Plans: Proc. of the 1986 Workshop*, Timberline, OR, June 30-July 2, 1987. Morgan Kaufmann Publ., Los Altos, CA.

[Green, 1969] C. Green, "Application of theorem proving to problem solving," In *Proc. of the Int. Joint Conf. on AI (IJCAI '69)*, pages 219–239, Washington, D. C., May 7-9, 1969.

[Haas, 1987] A.R. Haas, "The case for domain-specific frame axioms," In *F. M. Brown (1987)*, pages 343–348. 1987.

[Hanks and McDermott, 1987] S. Hanks and D. McDermott, "Nonmonotonic logic and temporal projection," *Artificial Intelligence*, 33:379–412, 1987.

[Hayes, 1987] P.J. Hayes, "What the frame problem is and isn't," In *Z.W. Pylyshyn (1987)*, pages 123–137. 1987.

[Kautz and Allen, 1986] H.A. Kautz and J.F. Allen, "Generalized plan recognition," In *Proc. of the 5th Nat. Conf. on AI (AAAI 86)*, pages 32–37, Philadelphia, PA, August 11-15, 1986.

[Kowalski, 1979] R.A. Kowalski, *Logic for Problem Solving*, volume 7 of *Artificial Intelligence Series*, Elsevier North Holland, New York, 1979.

[Kowalski, 1986] R.A. Kowalski, "Database updates in the event calculus," Technical Report DOC 86/12, Dept. of Computing, Imperial College, London, England, July 1986, 29 pages.

[Kowalski and Sergot, 1986] R.A. Kowalski and M.J. Sergot, "A logic-based calculus of events," *New Generation Computing*, 4:67–95, 1986.

[Kyburg, 1988] H. Kyburg, "Probabilistic inference and probabilistic reasoning," In Shachter and Levitt, editors, *The Fourth Workshop on Uncertainty in Artif. Intell.*, pages 237–244. 1988.

[Lansky, 1987] A.L. Lansky, "A representation of parallel activity based on events, structure, and causality," In *M.P. Georgeff and A.L. Lansky (1987)*, pages 123–159. 1987.

[Lifschitz, 1987] V. Lifschitz, "Formal theories of action," In *F. M. Brown (1987)*, pages 35–57. 1987.

[Manna and Waldinger, 1987] Z. Manna and R. Waldinger, "A theory of plans," In *M.P. Georgeff and A.L. Lansky*, pages 11–45. 1987.

[McCarthy, 1968] J. McCarthy, "Programs with common sense," In M. Minsky, editor, *Semantic Information Processing*, pages 403–417. MIT Press, Cambridge, MA, 1968.

[McCarthy, 1979] J. McCarthy, "First-order theories of individual concepts and propositions," In D. Michie, editor, *Machine Intelligence*, volume 9, pages 463–502. Edinburgh Univ. Press, Edinburgh, Scotland, 1979.

[McCarthy, 1980] J. McCarthy, "Circumscription – a form of non-monotonic reasoning," *Artificial Intelligence*, 13:27–39, 1980.

[McCarthy, 1984] J. McCarthy, "Applications of circumscription to formalizing commonsense knowledge," In *Proc. of the Nonmonotonic Reasoning Workshop*, pages 295–324, Menlo Park, CA, Oct. 17-19, 1984. Sponsored by AAAI.

[McCarthy and Hayes, 1969] J. McCarthy and P.J. Hayes, "Some philosophical problems from the standpoint of artificial intelligence," In B. Meltzer and D. Michie, editors, *Machine Intelligence*, volume 4, pages 463–502. Edinburgh Univ. Press, Edinburgh, Scotland, 1969.

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

[Morgenstern, 1987] L. Morgenstern, "Knowledge preconditions for actions and plans," In *Proc. of the 10th Int. Conf. on AI (IJCAI 87)*, pages 867–874, Milan, Italy, August 23-28, 1987.

[Morgenstern, 1988] L. Morgenstern, "Why things go wrong: a formal theory of causal reasoning," In *Proc. of the 7th Nat. Conf. on AI (AAAI 88)*, pages 518–523, Saint Paul, MN, August 21-26, 1988.

[Pearl, 1988] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufman, San Mateo, CA, 1988.

[Pylyshyn, 1987] Z.W. Pylyshyn, editor, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Ablex Publ., Norwood, NJ, 1987.

[Raphael, 1971] B. Raphael, "The frame problem in problem solving systems," In N. V. Findler and B. Meltzer, editors, *Artif. Intell. and Heuristic Programming*, pages 159–169. Edinburgh Univ. Press, Edinburgh, Scotland, 1971.

[Reiter, 1980] R. Reiter, "A logic for default reasoning," *Artificial Intelligence*, 13:81–132, 1980.

[Schank and Abelson, 1977] R.C. Schank and R.P. Abelson, *Scripts, Plans, Goals and Understanding*, Lawrence Erlbaum Assoc., Hillsdale, NJ, 1977.

[Weber, 1989] J. Weber, "Statistical inference and causal reasoning," In *Proc. 11th Int. Joint Conf. on AI (IJCAI '89)*, Detroit, MI, Aug. 20-25, 1989.