This is the version reprinted (with minor changes) in H. J. Levesque and F. Pirri (eds.), *Logical Foundations for Cognitive Agents*, Springer, Berlin, 1999, pp. 325-351. The original version is in *J. Logic and Computation 4*(5), pp. 679-700, 1994.

Explanation Closure, Action Closure, and the Sandewall Test Suite for Reasoning about Change

Lenhart K. Schubert

Department of Computer Science, University of Rochester
Rochester, NY 14627-0226

Abstract. *Explanation closure* (EC) axioms were previously introduced as a means of solving the frame problem. This paper provides a thorough demonstration of the power of EC combined with *action closure* (AC) for reasoning about dynamic worlds, by way of Sandewall's test suite of 12-or-so problems [29–31]. Sandewall's problems range from the "Yale turkey shoot" (and variants) to the "stuffy room" problem, and were intended as a test and challenge for nonmonotonic logics of action. The EC/AC-based solutions for the most part do not resort to nonmonotonic reasoning at all, yet yield the intuitively warranted inferences in a direct, transparent fashion. While there are good reasons for ultimately employing nonmonotonic or probabilistic logics – e.g., pervasive uncertainty and the qualification problem – this does show that the scope of monotonic methods has been underestimated. Subsidiary purposes of the paper are to clarify the intuitive status of EC axioms in relation to action effect axioms; and to show how EC, previously formulated within the situation calculus, can be applied within the framework of a temporal logic similar to Sandewall's "discrete fluent logic", with some gains in clarity.

1 Introduction

Explanation closure (EC) axioms are complementary to effect axioms. For instance, just as we can introduce effect axioms stating that painting or wallpapering a wall (with appropriate preconditions) changes its color, we can also introduce an EC axiom stating that a change in wall color implies that it was painted or wallpapered. The "closure" terminology signifies that the alternatives given are exhaustive.

This complementarity extends to their use: effect axioms allow the inference of change, and EC axioms the inference of non-change (persistence). For instance, if I know that no-one has painted or wallpapered the wall, then I can conclude that its color has remained unaltered. As first noted by Haas [15], EC-based persistence reasoning provides a very good handle on the *frame problem*.[1] In [32] (henceforth Sch90) I extended Haas' work, showing that EC-based techniques generalize to worlds with continuous and agentless

[1] A number of other writers have made closely related proposals, e.g., Lansky [20], Georgeff [9], Morgenstern & Stein [27].

change and concurrent actions, and support extremely efficient STRIPS-like methods for tracking effects of successive actions. Moreover, these methods are entirely monotonic as long as we are only concerned with inference of those changes and those explanations for change that are a matter of "practical certainty" (given our theory of the domain).

In view of their potency, it is surprising that EC-based approaches did not surface much sooner in the history of the frame problem. A commonly expressed qualm about EC axioms is that any enrichment of the (micro)world under consideration is likely to necessitate their revision. For instance, while in a simple world a change in wall color may be attributable to painting or wallpapering, in a more complex world the change may also be due to spraying, tiling, or panelling (or even decay, etc.). True enough – but it is equally true that enrichment of a microworld complicates the effect axioms. For instance, having paint and a brush may be sufficient for successful wall-painting in a simple, benign world, but in a more realistic one, the painter may be thwarted by dried-out paint, an undersize or oversize brush, injury, interference by other agents, etc. (i.e., the *qualification problem* crops up). Yet the fallibility of simple effect axioms has deterred few – not even nonmonotonic theorists – from relying on them! For instance, most formalizations of the Yale Turkey Shoot include axioms asserting that loading a gun makes it loaded, and firing the loaded gun at Fred kills him. This is generally done without comment or apology (except perhaps for a perfunctory gesture toward the qualification problem, which is thereafter ignored).[2] Yet the idea of turning this around and applying the same strategy to inference of explanations, given a change, seems to occur to almost no-one, and if raised, is met with skepticism.

I am led to believe that there are deep-seated prejudices against the idea of *reasoning deductively against the causal arrow*, perhaps stemming in part from the philosophical tradition on explanation. This tradition holds that physical theories enable us to deduce *resultant* states and events from given ones; while going from results to their causes is not a matter of deduction, but a matter of generating assumptions *from* which we can deduce the results. But while reasoning against the arrow of time and causation (retrodiction, explanation) is apt to generate more alternatives than reasoning with it (prediction), there is no *a priori* physical or logical reason for confining deduction to the forward direction.[3]

[2] A notable exception is [21], which explicitly addresses the qualification problem through circumscriptive minimization of preconditions. Also [13] addresses the qualification problem via a "possible worlds approach" (see the "stuffy room" scenario below).

[3] It is interesting that people versed in formal logic are apt to regard Sherlock Holmes' "deductions" as misnamed. Rather, they say, Holmes was reasoning inductively or abductively when he constructed explanations for his observations. In my view, if we are willing to grant that the inference of a man's death is deductive, given his unimpeded fall to the pavement from the top of a skyscraper, then

But the best argument against these prejudices lies in the practical efficacy of EC reasoning, of which there is growing awareness (as evidenced not only in [15], Sch90 and herein, but also in [5], [7], and [16]). Here we should also take note of an elegant and useful extension of the EC-based approach to the frame problem developed by Reiter [28]. Rather like Morgenstern and Stein [27], he focuses on cases where the known effect axioms characterize *all* the ways the changes of interest can come about (Generalized Completeness Assumption). For such cases, he shows how EC axioms can be derived from effect axioms, and combined with them into biconditionals ("successor state axioms"); e.g., a wall changes color *if and only if* it is painted or wallpapered. This mechanical derivation should allay some of the above qualms about the lack of invariance of EC axioms when new actions are added. Reiter further shows how to use such axioms for sound and complete goal regression.[4]

However, I will keep effect axioms and EC axioms separate for the sake of generality, since I believe that the GCA is valid only to the extent that the "blanket closure" assumptions implicit in nonmonotonic approaches are valid. It breaks down for realistically complex domains, and even for some simple worlds of interest. For instance, we may know that a robot's $Goto(x)$ action brings about $nextto(Robot, x)$. But it would be wrong to biconditionalize this to say that $nextto(Robot, x)$ becomes true *if and only if* the robot moves next to x. After all, there may be objects near x which the robot may also end up next to (and these "side effects" may depend more or less unpredictably on low-level path planning). Yet we can state an EC axiom that $nextto(Robot, x)$ becomes true *only if* the robot goes to some y and $x = y$ or x is near y; this may be quite sufficient for the persistence reasoning needed for practical purposes (see further details in Sch90). Sandewall's test suite provides additional illustrations [29,30] (henceforth San91, San92).[5] For instance, in the "stuffy room" problem (discussed at length later on), various EC axioms are possible (without change to the effect axioms), depending on how much freedom to "flit about" we want to allow objects when a vent is blocked or unblocked (creating drafts, one imagines). In general, we cannot characterize changes in terms of conditions that are both necessary and sufficient for those changes to occur. When we abstract away details in high-level axiomatizations (e.g., by using predicates like *nextto*), or have only partial knowledge of the behavior of a domain (because of its lack of familiarity, complexity, or inherent nondeterminism), then the best we can do is to provide some (practically certain) postulates about sufficient conditions for change, and others about necessary ones.

some of Holmes' inferences are equally deductive. If the former is not deductive, then no inferences based on world knowledge are deductive, whether directed forward or backward in time.

4 The usefulness of EC axioms in planning has also become apparent in more recent work on SAT-planning (e.g., [14]).

5 These publications were precursors of the monograph *Features and Fluents*[31].

The test suite provides an unprecedented opportunity to examine the strengths and shortcomings of various methods for reasoning about change in a systematic way. I will show that the approach based on EC-reasoning fares very well indeed. Moreover, the proffered solutions are monotonic except in the case of one variant of McCarthy's "potato in the tailpipe" problem (where I suggest a probabilistic approach). This seems to me to call for a reassessment of the proper roles of monotonic and nonmonotonic (or probabilistic) methods in reasoning about change. While nonmonotonic methods still retain an important role in reasoning about an uncertain, incompletely known world (as the "potato in the tailpipe" problem and other instances of the qualification problem show), monotonic methods can deal straightforwardly with many of the scenarios viewed as motivating examples for nonmonotonic methods.

The examples will also serve to illustrate a version of EC-based reasoning within a temporal calculus loosely modelled on Sandewall's DFL (discrete fluent logic). They will further illustrate the form and importance of *action closure* (AC) axioms in the temporal calculus, and allow us to probe the limits of the monotonic approach.

2 DFL, TC, and the test scenarios

Sandewall's *discrete fluent logic* (DFL), outlined in a preliminary way in San91 and developed into several variants in San92 (see also [31]), offers a concise notation for time-dependent descriptions of dynamic worlds. A very interesting aspect of DFL is the theory of entailment, whose central idea is that an agent can view the world as inert, with all fluents retaining their values *except* when forced to undergo change by the agent's actions. (In the model theory, each action has associated with it certain "trajectories" of change for a finite number of fluents, for each state in which the action may be initiated. I will have some further remarks about this semantics later on.) Another idea Sandewall pursues is that actions can "occlude" the fluents they may affect, for the duration of the action; i.e., the values of occluded fluents cannot be presumed to persist. A model preference criterion may then be employed according to which less occluded models and those that postpone *transparent* (non-occluded) change are preferred.

Of particular interest for my present purposes is Sandewall's effort to identify and catalogue many of the defects of extant nonmonotonic logics, and provide old and new test problems which bring these defects to light. Sandewall's preliminary assessment in 1991 was that his study "... provides reasons for renewed disappointment. The situation in 1991 is only marginally different from the one in 1986 [the year of the Hanks & McDermott paper]... most of the 'most popular' approaches actually fail on the test scenarios." (*ibid.*: sec. 7). In the more recent work (San92), however, the emphasis is on viewing various NM logics as "tools", whose utility for various purposes can be assessed via Sandewall's inertia-world semantics.

The "temporal calculus" (TC) notation I will use emulates Sandewall's DFL syntax to facilitate comparisons. Thus it consists of the usual first-order syntax plus the following DFL-like temporal notation (but without involvement of occlusion): Truth of a formula φ at (moment of) time τ is written $[\tau]\varphi$, and truth at all times in $[\tau_1, \tau_2]$ is written $[\tau_1, \tau_2]\varphi$. Also $[\tau_1, \tau_2]\varphi := v$ means that $[\tau_1]\neg(\varphi = v)$ and $[\tau_2]\varphi = v$, i.e., the value of φ becomes v somewhere in the interval $[\tau_1, \tau_2]$. If φ is a formula, we use $\varphi = T$ and $\varphi = F$ equivalently with φ and $\neg\varphi$ respectively (as in DFL). As a semantic basis for the notation so far, an interpretation of the fluent predicates and functions is assumed to provide their extensions at each moment of time. (The time line could be taken to be discrete or the real line.) We will also use an *action* predicate *do*, where $[\tau_1, \tau_2]do(\alpha, \beta)$ is true or false of an agent α, action β and time *interval* $[\tau_1, \tau_2]$, viz., the interval over which the action takes place. An interpretation of TC is assumed to specify the extension of *do* at all time *intervals*, rather than at all times. A useful abbreviation will be

$$[\tau_1 .. \tau_2]do(\alpha, \beta),$$

which stands for

$$(\exists\tau_1')(\exists\tau_2')[\tau_1 \leq \tau_1' \leq \tau_2' \leq \tau_2] \wedge [\tau_1', \tau_2']do(\alpha, \beta),$$

i.e., $do(\alpha, \beta)$ happens *somewhere* between τ_1 and τ_2. Though I will mostly use temporally annotated formulas of the types described, "timeless" formulas (e.g., specifying entity types) are also useful. These can be equivalently thought of as true at all times, and clearly the following is a sound rule of inference, for ϕ any formula: $\frac{\phi}{[t_1, t_2]\phi}$.

TC solutions to the test problems are generally more perspicuous and concise than solutions in the situation calculus (SC). (See examples of the latter in Sch90.) This is mainly because the TC notation allows us to index states of affairs directly via time variables, instead of requiring us to index them via sequences of actions. However, the most interesting difference lies in the way the action closure (AC) assumption – that all relevant actions are known – is encoded. In SC versions, the assumption is implicit in the functional dependence of situations on actions. In TC versions, times (and hence fluent values at those times) are introduced independently of actions, and so the assumption of complete knowledge of relevant actions needs to be stated separately. It will typically (though not always) be represented by the "only if" part of an equivalence of form, "x did y from time t_1 to time t_2 iff (x, y, t_1, t_2) is one of the following tuples...". Such axioms will be called "action chronicles" (with apologies to those, including Sandewall, who have employed the term differently).

An important question here is whether AC assumptions are by their nature excessively strong. Does it not require God-like omniscience to know what all the relevant actions are that could have affected the fluents of interest? The answer is no, provided that we are only looking for *practical* certainty rather than *absolute* certainty. The *relevant* actions are often ones which occur in a very limited spatiotemporal domain, for instance in a certain room during a short time interval. We often have good reasons to believe that we know all the relevant actions within such confines. For example, if we are physically "on the scene of the action", we can often be sure that we are aware of all the relevant physical actions (e.g., which objects were painted or moved about) thanks to our perceptual and cognitive abilities; and when there are possible relevant actions beyond our purview, we are often well *aware* of just what those gaps in our knowledge are.

If we are simply being told a *story*, we can rely on the narrator to withhold nothing of relevance from us. The narrator will not neglect to mention that Joe unloaded the gun before pulling the trigger on Fred. As Amsterdam [1] argues, narrators are expected to tell their story in a way that puts the hearer/ reader on the scene (vicariously, through the narrator's perceptions), and this entails reporting everything of relevance that happened. To be sure, there are many qualifications to be made and subtleties to be explored here. But my point is that the source of closure in narration is the narrator, not the hearer or God. (Formally, Amsterdam assumes that no actions occurred other than those deducible from the narrative, or that could have transpired during explicitly reported lapses in the narrator's awareness. I will have further comments on Amsterdam's proposals later.)

If instead a scenario represents a plan of action, whose consequences are yet to be observed (once the plan is carried out), then clearly it is the planner's *intention* to shield the fluents of interest from capricious disturbances. If you *plan* to kill Fred by loading the gun, aiming at Fred, and pulling the trigger, you surely plan *not* to unload the gun before pulling the trigger. And if you plan to repaint the walls a certain color, you surely do not intend to let others meddle at will. Thus it is the planner who is the source of action closure. He may ensure closure, for instance, by arranging to be the only agent on the scene, or to have only co-agents who will do his bidding, or who at least can be relied on not to interfere. That is all that is needed to justify AC axioms. Moreover, it is an important advantage of the explicit AC approach that we can arbitrarily *delimit* the spatial and temporal locations, the agents, and the kinds of actions for which our action chronicles are complete. By contrast, NM logics generally have much stronger, universal completeness assumptions built into their semantics, and this can lead to bizarre and unexpected inferences for larger, non-transparent examples.

Of course, if we demand *absolute* reliability of our axioms, then God-like omniscience is indeed required; after all, even the most carefully insulated and controlled setting is subject to freak occurrences. But that is not an observation about EC or AC axioms in particular, but about *all* nonlogical axioms. Moreover, a monotonic approach to the inference of change and persistence does not preclude the addition of *belief revision* mechanisms, capable of retracting, amending, or adding to the beliefs which form the basis for these monotonic inferences. When I discover that the wall I painted blue turned green when it dried, I'll revise my effect axioms; and if I find that while my back was turned, a prankster who had been hiding in the closet repainted the wall red, I'll revise my action chronicle. But unless and until that happens, I may well be best off reasoning monotonically with "practically certain" axioms.

The test scenarios which follow adhere closely to Sandewall's formulations. Each scenario is described very briefly, the intended conclusions are indicated, and then the TC formalization is shown. Although detailed proofs exist in all cases, the style of reasoning used to reach the desired conclusions should be clear enough from just a few sample proofs here and there. (The reader might in particular look at the reasoning given for the Hiding Turkey Scenario (HTS).) I hope that the axiomatizations are sufficiently transparent to allow the reader to reconstruct the rest. The headers are worth paying close attention to; they encapsulate essential dimensions of variation among test cases, largely as identified by Sandewall – dimensions often difficult for any one nonmonotonic logic to measure up to simultaneously.

In all of the axiomatizations, names beginning with `obs`, `chr`, `eff`, `exp`, and `ineq` respectively are used for axioms describing observations at particular situations or times, action chronicles, effect axioms, explanation closure axioms, and inequality axioms. These names serve no theoretical purpose, only a mnemonic one (unlike DFL conventions). As in Sch90, constants and functions will start with an upper case letter and variables and predicates will be lower case. Top-level free variables are implicitly universally quantified (with maximal quantifier scope). The predicate u ("unequal") takes any number of arguments and asserts that they are pairwise distinct.

Prediction: Yale Turkey Shoot (YTS)

There are two truth-valued fluents, a (alive) and l (loaded). Initially the turkey is alive and the gun not loaded. The agent loads, waits and fires. Loading brings about l (from prior state $\neg l$ or l), and firing brings about $\neg a$ and $\neg l$ provided that l held prior to it. We wish to conclude that at the end of firing, $\neg a$ holds (the turkey is not alive).

I will slightly embellish the usual action repertoire to include `Unload`, `Spin`, and `Chopneck`, for illustration and for consistency with later variants. For simplicity `Chopneck` has been given no preconditions and the effect axiom for `Unload` has been omitted, since these actions play no role here.

obs1 $[0]a \wedge \neg l$
chr1 $[t_1, t_2]do(Joe, y) \Leftrightarrow (t_1, t_2, y) \in \{(4, 6, Load), (10, 12, Fire)\}$
eff1 $[t_1, t_2]do(Joe, Load) \Rightarrow [t_2]l$
eff2 $[t_1]l \wedge [t_1, t_2]do(Joe, Fire) \Rightarrow [t_2](\neg a \wedge \neg l)$

eff3 $[t_1, t_2]do(Joe, Chopneck) \Rightarrow [t_2]\neg a$
exp1 $[t_1, t_2]l := T \Rightarrow [t_1..t_2]do(Joe, Load) \vee [t_1..t_2]do(Joe, Spin)$
exp2 $[t_1, t_2]l := F \Rightarrow (\exists y \in \{Fire, Unload, Spin\})[t_1..t_2]do(Joe, y)$
exp3 $[t_1, t_2]a := F \Rightarrow (\exists t_1')[t_1 \leq t_1' \leq t_2 \wedge [t_1']l \wedge [t_1'..t_2]do(Joe, Fire)]$
$\vee [t_1..t_2]do(Joe, Chopneck)$
ineq1 $u(Load, Unload, Fire, Spin, Chopneck)$

Reasoning: We infer $[4]\neg l$ by noting $[0]\neg l$ and that if $[4]l$ were true, a `Load` or `Spin` action would have had to occur between times 0 and 4, by `exp1`. But this is ruled out by `chr1`. Hence by `chr1` and `eff1`, $[6]l$. Similarly $[10]a$ since if this were false we would have had a `Fire` or `Chopneck` action between times 0 and 10 by `exp3`, contrary to `chr1` and `ineq1`.

Now we infer $[10]l$ in much the same way, using the fact that its falsity would imply a `Fire, Unload`, or `Spin` action by `exp2`, which can be ruled out by `chr1`. Hence by `chr1` & `eff2`, $[12](\neg a \wedge \neg l)$. $\neg l$ is easily shown to persist. $\neg a$ will persist if we add $[t](\neg a \wedge d \geq 0) \Rightarrow [t + d]\neg a$. □

Though superficially close to Sandewall's axiomatization, the TC version makes significantly stronger assumptions at the outset. For instance, `chr1` leaves Joe inactive between loading and firing, and this together with `exp2` ensures that the gun remains loaded. But in the DFL version, this is a defeasible chronicle completion inference. Should it be? Suppose the problem specification included the statement, "Between loading and firing, another action either did or did not take place". Intuitively, this blocks the inference that the gun remained loaded – despite the fact that the added statement is logically vacuous (a *tautology*)!

Clearly, it is a mistake to simply render the given English sentences as directly as possible in some logic, and then make it a matter of the semantics of that logic to deliver the intuitively required conclusions. How could *any* reasonable logic have entailments defeasible by tautologies? This once again raises the important question of "what's in a problem statement". As noted earlier, Amsterdam [1] drew attention to the role of narrative conventions in story-like problem statements, in particular the requirement that the author relate everything his audience would have observed under the reported circumstances – except perhaps events that transpired during explicitly reported lapses of attention (e.g., where the author indicates that some time passed, or says "I blacked out for a moment", etc.). This is formally written as UA_t, i.e., it is unknown whether action A occurred.

It is interesting to note that Amsterdam's assumption about what actions did and did not occur is closely related to the AC assumption. Stated a little more fully than before, his assumption is that an action A occurred at t if A_t is provable, and did not occur if neither A_t nor UA_t is provable. My AC assumption is computationally less problematic: it says that all the actions that bear on the fluents of interest are explicitly known, without invoking provability. Also, Amsterdam makes an assumption closely related to EC: roughly speaking, changes that are provable effects of provable actions (according to some theory of what constitutes an "effect") definitely occurred, and no change occurred unless it is the effect of some A_t, where A_t or UA_t is provable. (For the exact formulation, see [1].) Amsterdam notes that his approach fails to allow for actions which people regard as "obvious" inferences from certain state changes. His example is one where a character is sitting by the fireplace in one sentence and standing by the door in the next. These are precisely the action inferences supplied by EC!

Amsterdam's attempt to capture narrative conventions by nonmonotonic action and effect closure and the modal U operator is interesting, but it remains to be seen how far it can be taken. Besides computational intractability and the problem about action inference noted by Amsterdam, there is also the problem that real stories allow for many actions and events that are neither entailed by the story nor occluded by lapses in the narrator's attention. For instance, it certainly seems possible in a story like *Little Red Riding Hood* that the heroine hopped over a small creek, or glanced at some birds overhead on her way to Grandmother's house, even though nothing in the story entails this or even suggests that this *may* have occurred. The narrator simply did not judge such events relevant, and therefore, abiding by the Gricean maxim, omitted them. The view taken here is that narrative implicatures and domain reasoning are separable phenomena, and that it is therefore worthwhile to study domain reasoning methods as far as possible independently of story understanding. This means that we begin by extracting *all* of the information intuitively conveyed by a narrative – the positive as well as the negative, the asserted as well as the "conversationally implicated" information – while setting aside the question of exactly *how* the narration managed to convey that information. Only then do we ask what *follows* from what we have been told.

Regardless of strategy, however, what is important about Amsterdam's work is its recognition of the importance of narrative conventions and maxims in shaping what we take a story to imply. Much of the heated debate about which nonmonotonic logic is the right one for chronicle completion seems attributable to the neglect of information implicitly conveyed through these conventions and maxims, or misguided attempts to make this information fall out of the logic.

Retrodiction: the Stanford Murder Mystery (SMM)

The world is the same as for the YTS, but the gun is initially loaded, firing and waiting are performed in succession, and then the turkey is not alive. We are to infer that the gun was initially loaded, and the turkey was not alive after firing (prior to the wait).

obs1 $[0]a$
chr1 $[t_1, t_2]do(Joe, y) \Leftrightarrow (t_1, t_2, y) = (10, 12, Fire)$
obs2 $[14]\neg a$

eff1–ineq1 as above (YTS)

Ambiguous prediction: the Ferryboat Connection Problem (FCP)

A motorcycle M goes from F, some location on island Fyen, to the ferry landing L, and gets there between times 99 and 101. If it gets there before time 100, it will catch the ferry and be in Jutland (J) as of time 110, otherwise it stays at L. We are to infer that at time 110, M is either on L or on J (but should not infer one or the other).

Actually, Sandewall's DFL formalization makes the problem a little harder by saying, in effect:

> At time 0, the bike is on Fyen. At some time T between 99 and 101, the bike arrives at the landing. If its arrival T is before time 100, then the bike gets on board the ferry at time 100. If the bike is on board at time 105, it arrives on Jutland at time 110.

I will use a similar encoding for the TC version. The TC version assumes more, but, I will argue, rightly so.

obs1 $[0]on(M,F)$
chr1 $[t_1,t_2]do(M,y) \Leftrightarrow [(t_1,t_2,y)=(0,T,GotoL)] \vee [T<100 \wedge (t_1,t_2,y)=(100,101,Board)] \vee [T\geq 100 \wedge T\leq t_1\leq t_2\leq 110 \wedge y=Wait]$
obs2 $99\leq T\leq 101$
eff1 $[t_1,t_2]do(M,GotoL) \Rightarrow [t_2]on(M,L)$
eff2 $[t_1,t_2]do(M,Board) \Rightarrow [t_2]on(M,B)$
eff3 $[105]on(M,B) \Rightarrow [110]on(M,J)$
eff4 $[t_1,t_2]do(M,Wait)\wedge [t_1]on(M,y) \Rightarrow [t_1,t_2]on(M,y)$
exp1 $[t_1,t_2]on(M,B) := F \Rightarrow [t_1..t_2]do(M,Unboard)$
ineq1 $u(GotoL,Board,Unboard,Wait)$

Reasoning: Suppose $T<100$. Then by `chr1`, $[100,101]do(M,Board)$ and hence by `eff2`, $[101]on(M,B)$. By `exp1`, if $[105]\neg on(M,B)$ then $[101..105]do(M,Unboard)$, which can be seen to be false from `chr1`(according to which there are no actions beginning at time 101 or later if $T<100$). Hence $[105]on(M,B)$, and so by `eff3`, M gets to Jutland at time 110.

Now suppose $T\geq 100$. Then by `chr1` & `obs2`, $[T,110]do(M,Wait)$. (Likewise for all subintervals of $[T,110]$, but that doesn't interest us.) Also by `chr1`, $[0,T]do(M,GotoL)$ and hence by `eff1`, $[T]on(M,L)$. So by `eff4`, $[T,110]on(M,L)$.

Clearly there is no basis for supposing either $T<100$ or $T\geq 100$, and no unequivocal final location for M can be obtained. □

I said above that the TC version assumes more than Sandewall's DFL version. I was referring to the use of `Unboard` in the reasoning. The explanation closure axiom giving `Unboarding` as an explanation for $on(M,B)$ becoming false (i.e., `exp1`) is not just an embellishment but is essential to the inference that once M in on board, it stays on board till time 110.

I claim that this is an entirely reasonable assumption – in fact, that the desired conclusion about ending up at L or J should *not* be reached based on the information assumed by Sandewall. To illustrate the point, I will give a few syntactic variants of the "story" which lead to different conclusions.

At time 0, Albert is at home and hungry (state F).
At some time T between 99 and 101, Albert arrives (hungry) in the foyer of his favorite restaurant (state L).
If his arrival time T is before time 100, then he gets seated (still hungry) at time 100 (state B). (Maybe the manager usually holds a table for him till then, or maybe with the random comings and goings of customers it just happens to work out that way).
If he is seated and still hungry at time 105, he is seated and not hungry at time 110 (state J).

Can we conclude Albert is either in state L (hungry and in the foyer) or state J (seated and not hungry) at time 110? Clearly not: he could equally well be in state B (seated and still hungry), after having waited for a while in the foyer and finally gotten a table. He could even be in state F again (home and hungry) after stalking out of the restaurant, or even home and not hungry, having ordered and consumed a pizza (this is none of F, L, J). The point is that the conclusions we draw from even the simplest story about persistence of states are subtly dependent on world knowledge and narrative conventions, so we should not expect them to follow simply from superficial logical translations of the story sentences.

Here are two more variants:

At time 0, the subway is at station F.
At some time T between 99 and 101, it arrives at station L.
If its arrival T is before time 100, then it gets to station B at time 100.
If it is at station B at time 105, then it gets to station J at time 110.

In this case the alternative of still being at L seems quite unlikely. Moreover, the subway is not likely to be at J much past time 110.

At time 0, the house is on fire (but not wet) (state F).
At some time T between 99 and 101, the fire truck arrives and at that point the house becomes wet and on fire (state L).
If this happened before time 100 (say, the "flare point" of the fire), it will stop being on fire (while still being wet) at time 100 (state B).
If it was wet and not on fire at time 105, it'll be dry and not on fire at time 110. (state J).

Here again state B (wet and not on fire) can't be ruled out – the fire may have been doused by then anyway.

Prediction from disjunction: the Russian Turkey Shoot (RTS)

The problem differs from YTS only in that a `Spin` action (spinning the chamber of the gun) is inserted between the `Wait` and the `Fire`. The inference that the turkey dies should be disabled.

chr1 $[t_1, t_2]do(Joe, y) \Leftrightarrow (t_1, t_2, y) \in \{(4, 6, Load), (7, 9, Spin), (10, 12, Fire)\}$
obs1, eff1–3, exp1–3, ineq1 as in YTS

Ambiguous retrodiction: Stolen Car Problem (SCP)

At the beginning of the first night, the car is in my possession (expressed by predicate p). I perform the action of "leaving the car overnight in my garage" on two successive nights. On the following evening, the car is not in my possession.

I cannot lose possession of the car during the day. Once I've lost possession of it, I can't regain it. The intended conclusion is that I lost possession of the car during one of the two nights (with no conclusion about which night it was).

To illustrate that complete action closure is in general unnecessary, I will merely assume that the only `Leave-car-overnight` actions were those on the given nights ($[0, 2]$ and $[4, 6]$), so (given that only these can lead to car loss) the car couldn't have been lost during the day. The even weaker assumption that there were no `Leave-car-overnight` actions on the given days would have been sufficient, as well.

obs1 $[0]p$
chr1 $[t_1, t_2]do(I, Leave\text{-}car\text{-}overnight) \Rightarrow (t_1, t_2) \in \{(0, 2), (4, 6)\}$
obs2 $[8]\neg p$
eff1 $[t_1]\neg p \Rightarrow [t_1, t_2]\neg p$
exp1 $[t_1, t_2]p := F \Rightarrow [t_1..t_2]do(I, Leave\text{-}car\text{-}overnight)$

Random, but probable events: Ticketed Car Problem (TCP)

In some nonmonotonic approaches to the SCP above, the theft of the car would be treated as an exceptional event, and this will affect the axiomatization. Therefore San92 also offers a variant in which a car left overnight in a certain spot is *quite likely* to be ticketed. Other than that, the scenario and the desired conclusions are just as in the SCP. In the EC/AC approach, the distinction makes no difference so I omit the specifics.

Logically related fluents: Dead Xor Alive Problem (DXA)

This is a slight reformulation of the YTS, with "becoming not alive" replaced by "becoming dead", and the equivalence axiom $[t]\neg a \Leftrightarrow [t]d$ added (where

d means "dead"). Such logical connections lead to "autoramifications" (in Sandewall's terminology). In our monotonic approach the reformulation leads unproblematically to the conclusion that the turkey is d (and hence $\neg a$) after firing, and no sooner, much as before.

Logically related fluents: Walking Turkey Problem (WTP)

This is another slight variant of the YTS, in which the turkey is initially known to be walking (w) (but it is not explicitly given that he is alive), and the conditional $[t]w \Rightarrow [t]a$ is known. We are to conclude that the turkey is not walking after the firing. We easily infer $[0]a$ from $[0]w$ and reason as in YTS, concluding $[10]a$ and $[12]\neg a$ and hence $[12]\neg w$ by the contrapositive of the new conditional.

Prediction from disjunction: Hiding Turkey Scenario (HTS)

In this variant of Sandewall's, the turkey may or may not be deaf, and if it is not, it goes into hiding when the gun is loaded (where it is initially unhidden). Gun-loading, waiting, and firing take place in succession as in the YTS, but firing only kills the turkey if it is not hiding.

The intended conclusion is that at the end of firing, the turkey is either deaf and not alive, or nondeaf and alive. Sandewall points out that this problem confutes methods like Kautz's [19] which unconditionally prefer later changes to earlier ones (and so leave the turkey unhidden and hence deaf and doomed). In an EC-based approach, this variant is quite analogous to the RTS. We add an effect axiom that `Hide` brings about h (`eff4`), and EC axioms that *only* `Hide` and `Unhide` can bring about h and $\neg h$ respectively (`exp3`, `exp4`). We further add an assumption stating that if Fred is ever deaf, then he always was and always will be deaf (`eff5`).[6]

I will represent the gunman's (Joe's) and the turkey's (Fred's) actions by separate chronicles for clarity (`chr1` and `chr2`).

obs1 $[0]a \wedge \neg l \wedge \neg h$
chr1 $[t_1,t_2]do(Joe,y) \Leftrightarrow (t_1,t_2,y) \in \{(4,6,Load),(10,12,Fire)\}$
chr2 $[t_1,t_2]do(Fred,y) \Leftrightarrow [[5]\neg d \wedge (t_1,t_2,y) = (7,9,Hide)]$
eff1 $[t_1,t_2]do(Joe,Load) \Rightarrow [t_2]l$
eff2 $[t_1](l \wedge \neg h) \wedge [t_1,t_2]do(Joe,Fire) \Rightarrow [t_2](\neg a \wedge \neg l)$
eff3 $[t_1,t_2]do(Joe,Chopneck) \Rightarrow [t_2]\neg a$
eff4 $[t_1,t_2]do(Fred,Hide) \Rightarrow [t_2]h$
eff5 $[t_1]d \Rightarrow [t_2]d$
exp1 $[t_1,t_2]l := F \Rightarrow (\exists y \in \{Fire,Unload,Spin\})[t_1..t_2]do(Joe,y)$

[6] On a more careful analysis, the events causing or remedying deafness are like those causing or remedying a plugged car exhaust (see "Improbable disturbances" below). However, for the purposes of the present scenario it seems reasonable to treat deafness and nondeafness as permanent.

exp2 $[t_1,t_2]a := F \Rightarrow (\exists t'_1)[t_1 \leq t'_1 \leq t_2 \wedge [t'_1](l \wedge \neg h) \wedge [t'_1..t_2]do(Joe, Fire)]$
$\vee\ [t_1..t_2]do(Joe, Chopneck)$
exp3 $[t_1,t_2]h := T \Rightarrow [t_1..t_2]do(Fred, Hide)$
exp4 $[t_1,t_2]h := F \Rightarrow [t_1..t_2]do(Fred, Unhide)$
ineq1 $u(Load, Unload, Fire, Spin, Chopneck, Hide, Unhide)$

Reasoning: Suppose the turkey is initially deaf, $[0]d$. Then he is still deaf after the `Load` by `eff5`, hence he fails to `Hide` after the `Load` (or indeed, at any time) by `chr2`. Since he is initially unhidden according to `obs1`, he remains unhidden by `exp3`(etc.), so that in particular $[10]\neg h$. Likewise the l property inferrable at time 6 from the `Load` action and `eff1` persists by EC-reasoning to time 10. Hence the `Fire` action is fatal by `eff2`, and so $[12]\neg a$ and $[12]d$ (after another application of `eff5`).

On the other hand, if the turkey is not initially deaf, he is still nondeaf at time 5 during the `Load` (by the contrapositive of `eff5`). Hence Fred `Hides` during $[7,9]$ by `chr2`. He remains hidden through the subsequent actions by EC-reasoning based on `exp4`, in particular $[10]h$. After proving persistence of a from the initial state to time 10 in the usual way, we can also prove its persistence through the `Fire` action from `exp2`. Thus $[12]a$ and $[12]\neg d$ in this case (after another application of `eff5`). The assumption of initial deafness or non-deafness can each be made consistently, so that we can only infer the disjunction of the corresponding conclusions. □

Improbable disturbances: Potato in the Tailpipe (TPP)

Initially the car engine is not running ($\neg r$). The action of attempting to start the car is performed. On the assumption that there is usually no potato in the tailpipe (predicate p is usually false), and that the car will start if there isn't, we are to conclude that the car will start.

Sandewall approximates the premise that there is usually no potato in the tailpipe by saying that there is no potato in the tailpipe at time 0, by default. Default axioms are used only for final ranking of the most preferred models of the remaining axioms, and thus may be violated.

Ordinary first-order logic cannot express explicitly uncertain premises (such as ones involving "usually") and so cannot accurately model reasoning based upon them. To my mind the most attractive approach to uncertain reasoning is one based on *direct inference* of epistemic probabilities (i.e., probabilities for particular propositions) from "statistical" generalizations (e.g., [2], [3], [6], [17], [18]). The advantage of a probabilistic approach to nonmonotonicity is that it allows systematically for degrees of belief, and that it can provide a coherent basis for decision-making by an intelligent agent. The proof theory for direct inference is not yet fully developed, though the known techniques nicely handle many standard examples in nonmonotonic inheritance [6]. The present version of the TPP seems beyond the scope of current syntactic proof techniques, but can be analysed directly in terms of the model theory. For illustrations of direct inference for other variants of TPP, see [33], [34] and [3, 180-2].

The following, then, is a TC-like axiomatization of TPP based on a statistical interpretation of the statement about potatoes in the tailpipe. $[[t]\neg p]_t$ denotes the proportion of times t at which $[t]\neg p$ holds.

stat1	$[[t]\neg p]_t > .99$
obs1	$[0]\neg r$
chr1	$[t_1, t_2]do(Joe, y) \Leftrightarrow (t_1, t_2, y) = (6, 8, Start)$
eff1	$[t_1]\neg p \wedge [t_1, t_2]do(Joe, Start) \Rightarrow [t_2]r$

Reasoning: We appeal directly to the model-theoretic definition of epistemic probabilities [6]. Essentially $Prob([8]r|KB)$, where $KB =$ `stat1`$\wedge$ `obs1`$\wedge$ `chr1`$\wedge$ `eff1`, is the proportion of models of KB in which $[8]r$ holds. (More exactly, one considers the limit ratio as the number of time points comprising the time domain approaches ∞.)

Let the number of interpretations of p satisfying `stat1` be M (for some fixed finite time domain). Since in each of these interpretations the proportion of time points at which $\neg p$ holds is $> 99\%$, it is clear that more than 99% of these interpretations, say M' of them where $M' > .99M$, will have $[6]\neg p$ true in them.

Now each of these M' interpretations can be extended to a model of KB by using any interpretations of r and do that satisfy `obs1`$\wedge$ `chr1`$\wedge$ $[8]r$. (Note that this entire conjunction does not involve p.) So if there are N such interpretations of `obs1`$\wedge$ `chr1`$\wedge$ $[8]r$, we obtain $M'N$ models of KB in which $[6]\neg p$ and $[8]r$ hold.

This leaves the remaining $M - M'$ ($< .01M$) interpretations of `stat1` to be considered for which $[6]p$ holds. Each of these interpretations can be extended to a model of KB using any interpretations of r and do that satisfy `obs1`$\wedge$ `chr1` (since the antecedent of `eff1` will always be false when $[6]p$ holds, so that `eff1` will be satisfied regardless of the truth or falsity of $[8]r$). Of these interpretations of r and do, N satisfy `obs1`$\wedge$ `chr1`$\wedge$ $[8]r$, and (as is not hard to see) an equal number satisfy `obs1`$\wedge$ `chr1`$\wedge$ $[8]\neg r$.

Thus,

$$Prob([8]r|KB) = \frac{M'N+(M-M')N}{M'N+2(M-M')N} = \frac{1}{1+(M-M')/M} > \frac{1}{1.01} > .99. \ \square$$ [7]

As long as we demand that very improbable, but nevertheless possible, events be explicitly allowed for, the TPP cannot be monotonically represented. Still, the following trivial monotonic approximation is worth noting. Here the tailpipe is assumed to be initially clear, and the assumed chronicle and EC axiom for tailpipe plugging-up rule out any mischief.

[7] It is interesting to note that since we haven't said $[t_1]p \wedge [t_1, t_2]do(Joe, Start) \Rightarrow [t_2]\neg r$, the model counting technique predicts that with the tailpipe plugged, the car still has a 50% chance of starting, and that's why the lower bound on overall success probability is expected to be slightly better than 99/100, namely 100/101.

obs1	$[0]\neg r$
obs2	$[0]\neg p$
chr1	$[t_1, t_2]do(x, y) \Leftrightarrow (t_1, t_2, x, y) = (6, 8, Joe, Start)$
eff1	$[t_1]\neg p \wedge [t_1, t_2]do(Joe, Start) \Rightarrow [t_2]r$
exp1	$[t_1, t_2]p := T \Rightarrow (\exists x)[t_1..t_2]do(x, Plug)$

Event-time ambiguity: the Tailpipe Marauder (TPM)

This variant of Sandewall's assumes that a potato *is* put in the tailpipe somewhere between 8am and 5pm, but it is not known when. The attempt to start the car takes place at 1:30pm, and the aim is *not* to reach a conclusion about whether the car starts or not.

TPM is much less problematic for a monotonic approach than the original TPP, since it merely involves *incomplete* knowledge (about the time of a known event), rather than "defeasible" knowledge (where one of the possibilities consistent with our incomplete knowledge is much more probable than the others). I'll arbitrarily call the protagonist Joe and the antagonist Moe, and assign a duration of 2 (two hundredths of an hour) to the `Plug` and `Start` actions.

obs1	$[0]\neg r$
obs2	$[0]\neg p$
chr1	$[t_1, t_2]do(x, y) \Leftrightarrow (t_1, t_2, x, y) \in \{(1350, 1352, Joe, Start), (T, T{+}2, Moe, Plug)\}$, $800 \leq T \leq 1699$
eff1	$[t_1, t_2]do(x, Plug) \Rightarrow [t_2]p$
eff2	$[t_1, t_2]do(x, Start) \Rightarrow [[[t_1]p \Rightarrow [t_2]\neg r] \wedge [[t_1]\neg p \Rightarrow [t_2]r]]$
exp1	$[t_1, t_2]p := T \Rightarrow [t_1..t_2]do(Moe, Plug)$
exp2	$[t_1, t_2]p := F \Rightarrow [t_1..t_2]do(Joe, Unplug)$
exp3	$[t_1, t_2]r := T \Rightarrow [t_1..t_2]do(Joe, Start)$
ineq1	$u(Start, Plug, Unplug)$

It is straightforward to show that neither $[1352]\neg r$ nor $[1352]r$ can be inferred. Note that if we are *given* $[1352]\neg r$, we can infer $T < 1350$ and if we are *given* $[1352]r$, we can infer $T \geq 1350$.

Event-order ambiguity: Tailpipe Repairman Scenario (TPR)

In this variant, Sandewall assumes that the tailpipe is initially blocked, and the actions of unplugging the tailpipe and trying to start the car are done in arbitrary order. No action ordering should be inferrable, but it should follow that the car starts iff the unplugging is done first.

I include the gratuitous assumption that the tailpipe was unobstructed prior to 8am, for conformity with Sandewall's axiomatization.

obs1	$[800]\neg r$	(not running at 8am)
obs2	$[0]\neg p$	(no potato previous midnight)

obs3 $[800]p$ (potato in tailpipe at 8am)
chr1 $[[t_1, t_2]do(x, y) \wedge (800 \le t_1 \le 1698) \wedge (800 \le t_2 \le 1698)]$
$\Leftrightarrow (t_1, t_2, x, y) \in \{(T_1, T_1{+}2, Joe, Start), (T_2, T_2{+}2, Joe, Unplug)\}$
eff1 $[[t_1]\neg p \wedge [t_1, t_2]do(Joe, Start)] \Rightarrow [t_2]r$
eff2 $[t_1, t_2]do(Joe, Unplug) \Rightarrow [t_2]\neg p$
exp1 $[t_1, t_2]p := F \Rightarrow (\exists x)[t_1..t_2]do(x, Unplug)$
exp2 $[t_1, t_2]p := T \Rightarrow (\exists x)[t_1..t_2]do(x, Plug)$
exp3 $[t_1, t_2]r := T \Rightarrow (\exists t_1' \ge t_1)(\exists x)[t_1']\neg p \wedge [t_1'..t_2]do(x, Start)$
ineq1 $u(Start, Plug, Unplug)$

Neither $[T_1 + 2]r$ nor $[T_1 + 2]\neg r$ can be inferred. With assumption $T_2 + 2 \le T_1$, we would get $[T_1 + 2]r$, and for the contrary assumption we find the car will never run. Given the extra premise `obs2`, it is also possible to deduce a `Plug` action prior to 8am.

Conditional durations: Furniture Assembly Scenario

A furniture kit is initially unassembled. It is not known whether assembly instructions are included or not. (i or $\neg i$ may hold.) The `Assemble` action is performed, and this requires 20 minutes for completion if the instructions were included and 60 minutes otherwise. The desired conclusion is just that if the instructions were included, the kit is assembled within 20 minutes, and if not, within 60 minutes.

In the following axiomatization, T is the unknown assembly time. The chronicle says that the `Assemble` action is my only action, and could easily be refined to say it is my only action between times 0 and T. (In fact, we could just have asserted $[0, T]do(I, Assemble)$ – completeness is irrelevant here.) Inclusion of the instructions is treated as an atemporal (or permanent) property, though it could also be treated as a fluent. No EC axioms are needed and so none are shown.

obs1 $[0]\neg a$
chr1 $[t_1, t_2]do(I, x) \Leftrightarrow (t_1, t_2, x) = (0, T, Assemble)$
eff1 $i \wedge [t_1, t_2]do(I, Assemble) \Rightarrow [t_2]a \wedge t_2 = t_1 + 20$
eff2 $\neg i \wedge [t_1, t_2]do(I, Assemble) \Rightarrow [t_2]a \wedge t_2 = t_1 + 60$

Reasoning: We easily reach the conclusion that $[T]a$ and that $T = 20$ if i and $T = 60$ if $\neg i$, using reasoning by cases.□

A stable world: Lifschitz's N blocks

Lifschitz's N-blocks world [21] provides one example of a slightly more complex world than the previous ones. The world in the immediately following example (the "stuffy room" scenario) is likewise more complex, and also less sedate than the N-blocks world. As far as the EC/AC-based approach is concerned, neither presents any unusual challenge (and indeed at least equally complicated cases were treated in Sch90).

The N-blocks world allows movement of one block onto another (with the usual clear-top conditions, formulated in a slightly unusual way in terms of a *top* function) or onto the table, and painting of a block with one of three colors. There are axioms about uniqueness of destinations and resultant colors, and so on. The aim is to come up with the same state-transition characterization of this world as Lifschitz obtains circumscriptively, i.e., (roughly) nothing else changes when a block is moved or painted.

I will not spell out the details of the EC/AC-approach here, as this would be rather pointless. In essence, one just adds EC axioms about the 5 fluent predicates employed $(at, color, true, false, clear)$: blocks change *at* properties only when moved, and change color only when painted, etc. In fact, Reiter's technique for automatic biconditionalization of effect axioms would work well here. The reason it would be pointless to spell all this out is that in doing so, one would assert precisely what Lifschitz sets out to *prove* by circumscribing *causes* and *precond*!

The circumscriptive approach would be preferable if it could be depended on to give the desired persistence properties independently of the domain, without any need to specify EC axioms. Let me reiterate that this is not in general true, because we do not in general have complete knowledge of effects (recall *nextto*).

An unstable world: Agatha's stuffy room

Lifschitz's world is as stable as one would expect a blocks world to be, whereas in Ginsberg & Smith's "stuffy room" world [12] there is considerable latitude for objects to "flit about" unpredictably. They are apt to do so when Tyro, Aunt Agatha's robot, moves an object onto or away from one of the two ventilation ducts on the floor. This is claimed to be consistent with the intuition that light objects like newspapers may be shifted when the airflow in the room changes.

I will consider Winslett's variants of the original scenarios ([35]). These scenarios are designed to illustrate the advantages of Winslett's "possible models approach" (PMA) over Ginsberg & Smith's In essence, the advantage is insensitivity of nonmonotonic inferences to the syntactic form in which information is supplied.

Sandewall [30, 205-6] discusses the scenarios briefly, but does not attempt to duplicate the results in his DFL-2 logic. His reason is that he does not think the conclusions drawn are convincing. In particular, he questions the assumption that an object placed on a vent will stay put, while at the same time this blockage of air flow can cause motion of an object at *another* vent. More generally, he suggests that we should not equivocate about the causal model ("abstraction"): either things remain inert when we block a vent, or we should model the way in which blockage increases pressure, and the way in which this pressure in turn shifts lightweight objects.

Sandewall has a point. In fact, close scrutiny of the examples reveals that the minimization of net change in the PMA (as in the PWA) has some peculiar effects. For instance, the two vents act as "object attractors" when both are initially blocked and one of the blocking objects is removed. The reason turns out to be that by attracting a blocking object, the unblocked duct can maintain its own "blocked" property and the "stuffy" property of the room! On the other hand, when one vent is already blocked and a blocking object is placed on the second vent, the first vent is apt to blow off its obstructing object, so as to maintain its "unblocked" property – and the room's nonstuffy property. More generally, it seems quite odd to claim that the PMA (or the PWA) somehow allows precisely for the physically plausible sets of alternative side effects induced by an action. There surely is no limit to the number and type of potential side effects. Once we have opened the door to drafts, why should we not also admit effects transmitted through attached strings, magnets, electrical conductors, etc.? These may have been cleverly hidden, and be no more apparent to the eye than the ventilator drafts; and their relative *improbability* is surely not something that can magically pop out of the *logic* we use.

Notwithstanding all that, it is of interest to encode this slightly bizarre world monotonically, as a test of the flexibility of the EC-based approach. After all, one *can* make up a physics story about why the vents attract and repel objects as predicted by the PMA. Once one makes these physical assumptions explicit through EC axioms, the charge of arbitrariness will no longer stick. With regard to Sandewall's specific objection, one can imagine that Tyro holds on to an object after moving it, until the gusts caused by the changes in duct blockage have settled down. Encoding a more causally coherent world such as Sandewall envisages would certainly be possible as well, indeed easier. EC axioms allow us to tailor the persistence knowledge to fit the physics, relieving us from trying to make the physics fall out of the semantics.

We begin with a set of timeless "laws" constraining Aunt Agatha's living room R. Everything is either a location or is *on* something. For one thing to be on another, the latter must be a location while the former must not. There are exactly two floor ducts D_1 and D_2. These and the floor ($Floor$) are the only *locations*. A thing can be *on* only one location and only the floor can have more than one thing on it. A duct is *blocked* iff something is on it. The room is stuffy iff both ducts are blocked. Symbolically,

law1 $location(x) \vee (\exists y) on(x,y)$
law2 $on(x,y) \Rightarrow [location(y) \wedge \neg location(x)]$
law3 $duct(x) \Leftrightarrow x \in \{D_1, D_2\}$
law4 $location(x) \Leftrightarrow [duct(x) \vee x = Floor]$
law5 $[on(x,y) \wedge on(x,z)] \Rightarrow y = z$
law6 $[on(x,y) \wedge on(z,y)] \Rightarrow [z = x \vee y = Floor]$
law7 $[duct(d) \wedge (\exists x) on(x,d)] \Leftrightarrow blocked(d)$

law8 $[blocked(D_1) \wedge blocked(D_2)] \Leftrightarrow stuffy(R)$

Winslett's first scenario was designed to illustrate the difficulties that the PWA encounters with the frame problem when some properties of the initial state are entailed by the axioms but not explicitly asserted. Agatha's TV, birdcage C and magazine M are on D_1, D_2, and $Floor$ respectively. There is also a newspaper N but nothing is specified about it (except, one presumes, that it is distinct from the other things in this world). Note that if N is not a location it must be on a location (`law1`), and since the ducts are occupied, it must in fact be on the floor. The only available action (performable by Tyro) is $Move(x,y)$, for which it is necessary that y is the floor, or nothing is *on* y, or x is already *on* y. Under these conditions the effect is that x is *on* y.

By a careful consideration of the possible models in all 3 of Winslett's scenarios, we find that the EC laws of this world should say the following. First, an object can flit spontaneously to a duct only if that duct is initially blocked and Tyro moves away a blocking object from either one of the ducts. (Model-theoretically, this can avert changes in "blocked" and "stuffy" properties.) Second, an object can flit spontaneously to the floor only if it is on a duct initially, the other duct is not blocked, and Tyro moves an object from the floor onto the other duct. (Model-theoretically, this can avert a change from a nonstuffy room to a stuffy one.)

Agatha now asks Tyro to move the TV to the floor. The desired conclusion is that the TV will be on the floor, while other objects may flit to one or the other duct, in conformity with one of Winslett's 6 models.

obs1 $[0](on(TV, D_1) \wedge on(C, D_2) \wedge on(M, Floor))$
chr1 $[t_1, t_2]do(Tyro, x) \Leftrightarrow (t_1, t_2, x) = (1, 2, Move(TV, Floor))$
eff1 $[[y = Floor \vee [t_1]\neg on(z, y) \vee [t_1]on(x, y)] \wedge [t_1, t_2]do(Tyro, Move(x, y)]$
$\Rightarrow [t_2]on(x, y)$
exp1 $[[t_1,t_2]on(x, y) := T \;\wedge\; y \in \{D_1, D_2\}] \Rightarrow [t_1..t_2]do(Tyro, Move(x, y))$
$\vee [(\exists t \geq t_1)[t]blocked(y)$
$\wedge (\exists x')[t][(on(x', D_1) \vee on(x', D_2))$
$\wedge (\exists y')[t][\neg on(x', y') \wedge [t..t_2]do(Tyro, Move(x', y'))]]]$
exp2 $[[t_1, t_2]on(x, Floor) := T \Rightarrow [t_1..t_2]do(Tyro, Move(x, Floor))$
$\vee [(\exists t \geq t_1)(\exists d_1, d_2 \in \{D_1, D_2\})[t]on(x, d_1) \wedge \neg on(z, d_2)$
$\wedge [t..t_2]do(Tyro, Move(x', d_2))]$
ineq1 $u(D_1, D_2, Floor, M, N, TV)$
ineq2 $(x \neq x' \vee y \neq y') \Rightarrow u(Move(x, y), Move(x', y'))$

Reasoning: First, we obviously get $[2]on(TV, Floor)$ from `chr1` and `eff1`. This persists to time 3 since if it became false, $on(TV, x)$ would have to become true for some x and so by `exp1` there would have to be an additional `Move` between times 2 and 3, contrary to `chr1`. The additional conclusions desired can be reformulated as follows:

1. If nothing flits to duct D_1, then only the TV moves to a new location;

2. If one of M, N flits to D_1, then only that object and the TV move to a new location;
3. If C flits to D_1, then there are no constraints on further flitting except those dictated by the "laws". (So there is nothing further to be proved in this case.)

First we note that when Tyro moves the TV to the $Floor$, he makes no other concurrent move. This follows from `chr1`, `ineq1` and `ineq2`. So any additional shifts are due to "flitting". So to prove (1), assume nothing flits to D_1, so that $[2]\neg on(x, D_1)$. To complete the analysis for time 2, we need only show that cage C does not flit to the $Floor$. If C did flit to the $Floor$, the second disjunct of `exp2` would apply, and `chr1` would force the identification $t = 1$; at that time, if $d_1 = D_1$ then $on(x, d_1)$ would be false, and if $d_2 = D_1$ then $\neg on(z, d_2)$ would be false. In either case `exp1` would be violated and so C does not flit to the $Floor$.

To prove (2), assume first that magazine M flits to D_1. Once again we need only show that C does not flit to the floor. As before we find that if it did, the second disjunct of `exp2` would be violated. The argument for the case that newspaper N flits to D_1 is completely analogous. It remains to show that there is no change from time 2 to time 3, and this follows from the fact that `exp1` would require a further action by Tyro between those times for any *on* relationship to change, and this is ruled out by `chr1`. (The "laws" then also prevent change of *blocked* and *stuffy*.) □

Second and third "stuffy room" variants

Winslett's second variant was designed to show that the PWA can generate unwarranted conclusions in the presence of a logically redundant disjunction, and the third to show that it makes a difference for inferences under the PWA whether or not an entailed negative literal is explicitly present. I will not go into these except to say that for the effect and EC axioms above one obtains the same alternative outcomes for these scenarios as are obtained by Winslett's PMA.

Concurrent actions

In Sch90 I suggested that many of the alleged deficiencies of the situation calculus, as a general calculus for action and change, were due simply to neglect of the possibilities inherent in *functions* of situations and actions. In particular, I suggested that (1) external change could be accommodated by letting the usual $Result(a, s)$ function predict such change (for instance, *Result*(*Wait-a-minute,s*) might differ significantly from s if s is a dynamic situation such as one where the sun is about to rise); (2) continuous time and continuous change could be accommodated with functions like *Clock-time*(s) and $Trunc(a, t)$, where the latter supplies an initial segment of duration t

of action a (so that $s' = Result(Trunc(a,t), s)$ is a situation t seconds after situation s and $Clock\text{-}time(s') = Clock\text{-}time(s) + t$); and (3) most importantly composite actions, including concurrent ones, could be accommodated through action composition functions such as $Seq(a,b)$ and $Costart(a,b)$.

I worked through an example involving a man, a robot, and a cat, where the man walks from one place to another while the robot concurrently picks up and carries a box containing the (inactive) cat. I showed how persistence reasoning based on EC could be extended to such a setting. For instance, the EC axiom for color change says that if the color of an object is changed in the result state of a composite action, then that action must have had a primitive part in which the object is painted or dyed. I also showed how the usual effects of independent actions executed concurrently could be predicted if actions are provably *compatible*. In the example, compatibility of the concurrent actions was taken to be a consequence of disjointness of the "action corridors" within which they happen to occur. Rather than repeating such an example here, let me just reiterate that the solution was entirely monotonic, and that it can easily be recast in TC form.

Gelfond *et al.*[25] independently made some proposals similar to my own concerning the use of action combinators for dealing with concurrent actions in the SC. Just as I was concerned with showing certain concurrent actions to be *compatible*, they are concerned with showing that composite actions (with concurrent components) are *free of conflict*. For them, this means that the concurrent components do not have effects leading to different values for the same fluent, and they employ circumscription of conflict to minimize this sort of adverse interaction. While I considered only the case where compatibility (lack of conflict) is due to action disjointness, Gelfond *et al.* also allow for *constructive* interference, whereby certain effects of individual actions (like spilling of soup in one-handed lifting of a soup bowl) are "cancelled" in the concurrent case. These are interesting ideas, though their formulation is limited by the need to specify causal relations in a situation-independent way (*a la* [Lifschitz 1987]; *cf.,* [Baker 1991]). More importantly from the present perspective, the "blanket closure" assumptions implemented through circumscription are too strong, for much the same reasons that closure of effects is in general too strong. (Interference is, after all, due to the *effects* of actions on each other.)

Lin and Shoham [24] provide a third, and also closely related, preliminary proposal for allowing concurrency in SC. Their main concurrent combinator is written with set brackets, i.e., $\{\text{Action}_1, ..., \text{Action}_n\}$. Much as in the earlier attempts, a central concern is encoding noninterference between concurrent actions. They make the apt observation that this problem is analogous to the frame problem, i.e., by and large, actions don't interfere; accordingly, they tackle the problem by circumscriptively minimizing pairwise "cancellation" of given concurrent actions in given situations. In keeping with my approach to the frame problem, I would instead suggest the use of EC-like axioms to rule out interference; i.e., we specify various *necessary* conditions for various kinds of actions to interfere, and rule these out *where our observations and world knowledge allow us to do so.* In fact, the reasoning about "action corridors" in Sch90 I mentioned above involves just such an EC-like axiom, viz., one that states that for the physical motion of two objects to interfere, their paths must intersect – a reasonable postulate in many settings. If their paths are known *not* to intersect within a given time frame then we can infer noninterference. In this way we can avoid the extreme requirement of "epistemic completeness" which Lin and Shoham propose as a desideratum for action formalizations.

The ramification problem

The ramification problem arises from the fact that the changes directly produced by an action can entail additional changes, which can entail still further changes, and so on. The problem is to avoid exhaustively enumerating all the resultant changes in describing the effects of an action, yet be able to infer those changes (as needed).

Ginsberg [11] apparently regards the ramification problem as a difficulty for EC. However, a rather elaborate "robot's world" example in Sch90 [section 3] showed how well EC works even in the presence of ramifications. The ramifications in that example are ones resulting from arbitrarily stacked containment and *on*-relations. For instance, the robot may carry a box which contains a cup which in turn contains an egg. The inference that the cup and egg are transported along with the box is easily made. We merely need to be sure that the *in* and *on* relations persist, and for this we apply straightforward EC axioms stating, for example, what needs to happen in order for an *in*-relation to change. (In the axiomatization in Sch90, the robot needed to take the object out of the container, or take something else out of the container that "carries" the object along with it; "carries" was in turn axiomatized to allow for stacked in/on/part-of relations.)

I see no particular difficulties in extending these techniques to arbitrarily complex worlds. We neither need to *directly* axiomatize the cascaded effects of an action (rather we need only axiomatize "one link at a time" of the causal chains), nor retrace these cascades explicitly in the EC axioms.

3 Coda: The Metaphysics of Change

A recent trend in NMR research has been the development of criteria of *correctness* for nonmonotonic theories of action, based on "inertial" models. Sandewall [30,31] is a prime example of this, and Gelfond & Lifschitz [8] is in a similar spirit. In particular, these correctness criteria assume that (A) the world is totally inert except for changes wrought by the agent (Sandewall's "ego"), and (B) we have total knowledge of actions and action laws (and state constraints, if allowed).

I do not accept (A), i.e., "commonsense inertia". I think that while this is a reasonable working assumption for some highly restricted, tightly controlled domains, it is an untenable metaphysical position *vis a vis* "the world at large".

We do certainly *seek* invariants in the way we conceptualize the world; and to the extent we succeed, we reduce its perceived complexity and can cope more readily with it. But to extrapolate from our partial success in this quest for invariance to a metaphysics according to which the world really *is* a set of passive objects with static features, altered only by the intervention of one or a few agents, seems to me utterly implausible. Does anyone, naive or sophisticated, actually *believe* this? Ought not semantics accord with our intuitions about the world?

Perhaps academic researchers contemplating these matters are more apt than others to be impressed with the stability of the world, as their gaze wanders over tranquil office furnishings and inert papers and books, and their obedient workstation passively awaits the next keystroke. If in addition they understand computation, they may find the notion that the world is like a computer's internal store, modified only at the behest of the CPU, nearly irresistible (and indeed this analogy has been alluded to by the father of "commonsense inertia" – McCarthy [26]). I suspect that cab drivers, factory workers, weather reporters, stockbrokers, firemen or fishermen may have less affinity for such a metaphysics.

In my own metaphysics only the *laws* that govern fluents are constant, not the fluents themselves. The world is a chaotic place teeming with activity and change at all time scales and "granularities"; and only careful choice of vocabulary and coordinate frame and calculated neglect of many obvious variables imposes a semblance of order and stability on some patches of this hubbub.

But aren't these patches of stability enough to afford the proponents of inertia worlds a foothold? Yes, but note that this amounts to a pretense: it's not that the world *is* inert, just that for some purposes we can do business *as if* it were. And this pretense, like any other, is brittle: it lacks the robustness of truth, breaking down at the edges as we shift out of the narrow domain for which the pretense was contrived.

Of course, to the inertia adherents this simply indicates the need for a certain nimbleness in switching from one pretense to another – shifting to a new coordinate frame, a new vocabulary of fluents and actions reflecting new criteria of relevance, and semantically, to a new make-believe world of inert objects (see [22]). So there is one frame for physical action within the confines of the office, another for coping with rush-hour traffic, another for maintaining the lawn, others for functioning as part of various social groups and organizations (teaching, administering, parenting, choir singing, etc.), and so on.

But such nimbleness will be hard to achieve, since (i) the number of special domains is large, (ii) the boundaries are blurry, (iii) they can merge rather arbitrarily (a business meeting in an office, parenting while driving in rush-hour traffic, etc.), (iv) they don't always admit a static view of the relevant aspects, however adroitly we choose our frame, and (v) worst of all – at least from a logicist perspective – the *knowledge* of what frame is appropriate under what circumstances can have no coherent semantics, since there is no comprehensive inertia world in which the various make-believe micro-worlds can be embedded. The real world just *isn't* inert!

Wouldn't it be preferable to view the world realistically in the first place – exploiting the stabilities and regularities we find, of course, but treating these as contingent knowledge, for instance as knowledge about (commonsense) physics, or (commonsense) psychology, etc., rather than as a matter or *meta*physics? I claim that this can and should be done, through appropriate, limited explanation closure axioms (in conjunction with effect axioms).

Nor do I accept the epistemic assumption (B). Briefly, (i) We don't know all the actions in the world that have taken place or will take place. (ii) We don't know all the effects of all the actions we know about on all the fluents we care about. (iii) We don't know all relevant state constraints. (iv) We live in a world where there are many other agents as well as spontaneous change. (v) Effects may ramify unboundedly and affect unboundedly many fluents. (E.g., consider an object's distance from all others, when that object is moved.)

In short, it seems to me that methods based on inertia-world semantics are forever doomed to be applicable only to narrow, largely passive, insulated, thoroughly "predigested" domains, where moreover we have more or less complete knowledge of relevant actions and fluents, and the laws that govern them. The EC/AC approach lacks these metaphysical and epistemic overcommitments, yet can deal with the frame problem and has the flexibility to encompass multiple domains without inconsistency.

4 Conclusion

I have tried in this report to explore the scope of a particular technique, EC/AC-based reasoning in dynamic worlds, more fully than is the standard practice. I hope to have provided enough of the technical gist of the proposed EC/AC-based solutions to Sandewall's test suite to support my contention that much of the reasoning commonly thought to require nonmonotonic methods can in fact be done monotonically.

I should reiterate that in saying this, I am not suggesting that monotonic reasoning is all you really need. A monotonic theory of any realistically complex, dynamic world is bound to be an approximation, in the sense that it ignores both improbable qualifications on the effects of actions, and farfetched explanations for change. We simply cannot *express* in ordinary FOL that certain kinds of events are very unlikely but may nonetheless occur. For this, we need to go beyond FOL, as has been done in nonmonotonic and probabilistic logics.

But I think the literature on nonmonotonic logics has put *too much* of the burden of commonsense reasoning (especially too much of the task of inferring persistence and change) on nonmonotonic methods. An adverse effect has been a confusion between narrative principles and logic, and between physics and logic. The very terms "persistence" and "inertia" used as *model-theoretic* notions are suspect, since objects stay put, or keep moving, for physical rather than model-theoretic reasons. As well, the over-deployment of nonmonotonic methods has created computational intractability problems, where relatively simple monotonic methods would have sufficed.

The EC/AC-based approach seems to deal with most of the issues addressed by Sandewall's test suite rather handily. It does not render things quiescent (or nonexistent) merely because nothing is known about them, it does not spawn spurious events to minimize change, it does not fail when aimed backward in time, and it does not arbitrarily choose between disjuncts. Plausible EC and AC axioms are not hard to conjure up (and as Reiter showed, the former can sometimes be obtained mechanically), they do not work in mysterious ways, and they work computably and even efficiently (in STRIPS-like settings). It therefore seems well worthwhile to further investigate EC/AC-based methods, e.g., for planning applications. One of the most interesting directions for further work is to use probabilistically qualified EC and AC axioms in a probabilistic logic setting (cf. the earlier citations of work by Bacchus and Tenenberg & Weber); i.e., we would say such-and-such a change is *very likely* due to this or that kind of action, and such-and-such actions are *very probably* the only relevant ones that occurred in a certain setting. At that point we would be ready to address the qualification problem in full, while still exploiting the power of EC and AC to infer (probable) persistence or change.

Acknowledgements

Conversations with Ray Reiter and Andy Haas have clarified my understanding of the relation between effect axioms, EC axioms, and the qualification problem. The paper also benefited from Michael Georgeff's, Chung Hee Hwang's, Vladimir Lifschitz', and anonymous JLC referees' helpful comments. Support was provided by ONR/DARPA research contract no. N00014-82-K-0193 and Rome Lab contract F30602-91-C-0010. This paper previously appeared in *Journal of Logic and Computation 4*(5), pp. 679-799, 1994, and is reprinted with the permission of Springer-Verlag.

References

1. J. Amsterdam. Temporal reasoning and narrative conventions. In *Proc. of the 2nd RInt. Conf. on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 15–21, Cambridge, MA, April 22-25 1991.
2. F. Bacchus. Statistically founded degrees of belief. In *Proc. of the 7th Bienn. Conf. of the Can. Soc. for Computational Studies of Intelligence (CSCSI '88)*, pages 59–66, Edmonton, Alberta, June 6-10 1988.
3. F. Bacchus. *Representing and Reasoning with Probabilistic Knowledge*. MIT Press, Cambridge, MA, 1990.
4. 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.
5. E. Davis. Axiomatizing qualitative process theory. In *Proc. of the 3rd Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 177–188, 1992.
6. J.Y. Halpern F. Bacchus, A.J. Grove and D. Koller. Statistical foundations for default reasoning. In *Proc. of the Int. Joint Conf. on Artificial Intelligence (IJCAI-93)*, pages 563–9, 1993.
7. G. Ferguson and J.F. Allen. Actions and events in interval temporal logic. *Journal of Logic and Computation*, 4(5) (Special Issue on Actions and Processes):531–579, 1994.
8. M. Gelfond and V. Lifschitz. Representing actions in extended logic programming. In K. Apt, editor, *Proc. of the Joint Int. Conf. and Symp. on Logic Programming*, pages 558–573. MIT Press, 1992.
9. M.P. Georgeff. Actions, processes, causality. In M.P. Georgeff and A.L. Lansky (1987), pages 99–122, 1987.
10. 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.
11. M. Ginsberg. *Essentials of Artificial Intelligence*. Morgan Kaufmann, Los Altos, CA, 1993.
12. M. Ginsberg and D.E. Smith. Reasoning about actions i: A possible worlds approach. In F. M. Brown (1987), pages 233–258. 1987. Also in *Artificial Intelligence 35* (1988): 165–195.
13. M. Ginsberg and D.E. Smith. Reasoning about actions ii: The qualification problem. In F. M. Brown (1987), pages 259–287. 1987. Also in *Artificial Intelligence 35* (1988): 311–342.
14. D. McAllester H. Kautz and B. Selman. Encoding plans in propositional logic. In *Proc. of the 5th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'96)*, pages 374–384, Cambridge, MA, November 5-8 1996.
15. A.R. Haas. The case for domain-specific frame axioms. In F. M. Brown (1987), pages 343–348. 1987.
16. A.R. Haas. A reactive planner that uses explanation closure. In *Proc. of the 3rd Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 93–102, Cambridge, MA, 1992.
17. J.Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990.

18. Jr. H.E. Kyburg. Probabilistic inference and probabilistic reasoning. In Shachter and Levitt, editors, *The Fourth Workshop on Uncertainty in Artif. Intell.*, pages 237–244. 1988.
19. H. Kautz. The logic of persistence. In *Proc. of the 5th Nat. Conf. on AI (AAAI 86)*, pages 401–405, Philadelphia, PA, August 11-15 1986.
20. 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.
21. V. Lifschitz. Formal theories of action. In F. M. Brown (1987), pages 35–57. 1987.
22. V. Lifschitz. Frames in the space of situations. *Artificial Intelligence*, 46:365–376, 1990.
23. F. Lin and Y. Shoham. Provably correct theories of action (preliminary report). In *Proc. of the 9th Nat. Conf. on AI (AAAI-91)*, pages 349–354, Anaheim, CA, 1991.
24. F. Lin and Y. Shoham. Concurrent actions in the situation calculus. In *Proc. of the 10th Nat. Conf. on AI (AAAI-92)*, pages 580–585, San Jose, CA, 1992.
25. V. Lifschitz M. Gelfond and A. Rabinov. What are the limitations of the situation calculus? In *Working Notes, AAAI Symp. on Logical Formalizations of Commonsense Reasoning*, pages 59–69, Stanford Univ., Stanford, CA, 1991.
26. J. McCarthy. The frame problem today. In F. M. Brown (1987), page 3. 1987. (abstract).
27. L. Morgenstern and L.A. Stein. 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.
28. R. Reiter. The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation*, pages 359–380. Academic Press, 1991.
29. E. Sandewall. Features and fluents. Technical Report Res. Rep. LiTH-IDA-R-91-29, Dept. of Computer and Information Science, Linköping University, Linköping, Sweden, 1991. Review version of parts of a book.
30. E. Sandewall. Features and fluents. Technical Report Res. Rep. LiTH-IDA-R-92-30, Dept. of Computer and Information Science, Linköping University, Linköping, Sweden, 1992. Second review version of parts of a book.
31. E. Sandewall. *Features and Fluents. The Representation of Knowledge about Dynamical Systems. Volume I.* Oxford University Press, 1994.
32. L.K. Schubert. Monotonic solution of the frame problem in the situation calculus: an efficient method for worlds with fully specified actions. In R. Loui H. Kyburg and G. Carlson, editors, *Knowledge Representation and Defeasible Reasoning*, pages 23–67. 1990.
33. J. Tenenberg. Abandoning the completeness assumption: a statistical approach to solving the frame problem. *Int. J. of Expert Systems*, 3(4):383–408, 1990.
34. J. Tenenberg and J. Weber. A statistical approach to the qualification problem. Technical Report Technical Report 397, Dept. of Computer Science, Univ. of Rochester, Rochester, NY, 1992.
35. M. Winslett. Reasoning about action using a possible models approach. In *Proc. of the 7th Nat. Conf. on AI (AAAI-88)*, pages 89–93, Saint Paul, MN, August 21-26, 1988.