Minimizing the Effects of Actions

Tom Costello
Computer Science Department
Stanford University
Stanford, CA 94305
email: costello@cs.Stanford.EDU

Abstract

The frame problem—representing the default that the only effects of actions are those mentioned—is a central problem in knowledge representation. This default can be captured, using circumscription, by minimizing the set of *change formulas* that are true, where a change formula is a formula that states that, under certain circumstances, an event causes a fluent to change its value. This approach handles domain constraints and arbitrary disjunctions correctly, unlike previous proposals. It applies a single circumscription policy to the entire theory, rather than use ad-hoc methods that depend on applying several circumscriptions or splitting up the theory before circumscription is applied.

1 The Frame Problem

Reasoning about change is central to much of common sense reasoning. It is difficult to imagine how any entity could be called intelligent and be unable to reason about how things change.

To reason about change we clearly need to represent what changes after an action is completed. What is not so immediately obvious is that we also need to specify what does not change after an action. The simplest way to specify what changes and what does not would be to write each set of changes down. However this seems wasteful, and worse, seems likely to introduce inconsistencies. It seems clear that just specifying what changes is sufficient—everything else doesn't change. We can generate the properties that do not change from the list of properties that do change. If we do this we are guaranteed that our lists will be consistent. Moreover, usually fewer things change than don't change and thus we will have much less to specify.

The idea that we need specify just the positive instances of change has given rise to a problem: How do we generate all the instances of properties that stay the same? This problem is known as the *frame problem*.

1.1 The Situation Calculus

To look at the frame problem we need a logical language to represent time, events, and the properties that hold in the world. We propose to use the situation calculus, a formalism for reasoning about events in the world. The paper that introduced situations [MH69] defined them as follows, "A situation s is the complete state of the universe at an instant of time". The basic mechanism used in the calculus to define a new situation is the *result* function,

$$s' = result(e, s).$$

In this formula s is a situation, e is an event, and s' is a new situation that results when e occurs. A function or predicate of a situation is called a *fluent*.

1.2 Frame Axioms

As we mentioned earlier, a certain problem arises when we formalize domains in this language. Assume there is an axiom that states what happens to the location of a block when an action *move* is performed. If another property, *colour*, is added the axiom which described what effect the action had on the location does not give us any information on what happens to the new fluent *colour*. The usual intent is that the new fluent remains unchanged by the action, but this is not always the case. To specify that the colour of the block does not change we must write:

$$\forall x\ c\ l\ s. \mathrm{holds}(colour(x,c),s) \rightarrow \mathrm{holds}(colour(x,c),\mathrm{result}(move(x,l),s))$$

To express all the fluents that do not change an axiom would be needed for each action/fluent pair. These axioms stating that fluents do not change are called *frame axioms*. Having to specify all this lack of change seems extravagant and more importantly it seems unnecessary to have to write these out individually. It seems that it should be possible to generate these frame axioms automatically. The *frame problem* is the problem of characterizing these frame axioms.

1.3 Non-monotonic Logic

We cannot hope that classical logic will generate the frame axioms for us as classical logic is monotonic and will not infer any sentence that might be denied by an additional fact. Thus we will have to use a logic without this property, a non-monotonic logic.

Non-monotonic logics are logics which can make default assumptions which can be withdrawn later. We are looking at an assumption, "the tendency of facts to endure over time" [HM86]. This seems to be exactly the sort of problem at which non-monotonic reasoning should excel.

1.4 What a solution should generate

Ideally our non-monotonic logic should generate all of the frame axioms. If we look at an action that has preconditions such as

$$\forall x\ l\ s. \mathrm{holds}(\mathrm{movable}(x),s) \rightarrow \mathrm{holds}(\mathrm{at}(x,l),\mathrm{result}(\mathrm{move}(x,l),s))$$

we should get the frame axiom:

$$\begin{aligned}\forall x\ l\ s.\ & \neg\mathrm{holds}(\mathrm{movable}(x),s) \rightarrow \\ & (\mathrm{holds}(\mathrm{at}(x,l),s) \equiv \mathrm{holds}(\mathrm{at}(x,l),\mathrm{result}(\mathrm{move}(x,l),s)))\end{aligned}$$

We should have a frame axiom whenever there is no action whose preconditions are met that changes the fluent. We can write this as

> If there is no action, all of whose preconditions are met, that changes the fluent f then the value of the fluent f should remain the same.

This is a statement of the common sense law of inertia. If this is to generate enough frame axioms we must assume that actions have as few effects as possible. Therefore we must look at how we represent the effects of actions.

If an action a causes a fluent f to be false when the set of fluents P are all true and the fluents N are all false then we write:

$$\forall s. \bigwedge_{f_i \in P} \mathrm{holds}(f_i, s) \wedge \bigwedge_{f_j \in N} \neg \mathrm{holds}(f_j, s) \rightarrow \neg \mathrm{holds}(f, \mathrm{result}(a, s))$$

This does not tell us whether the action a changed f as f could have been false initially, i.e. in N. The sentence that tells us just the negative changes that the action a makes is:

$$\forall s. \bigwedge_{f_i \in P} \mathrm{holds}(f_i, s) \wedge \bigwedge_{f_j \in N} \neg \mathrm{holds}(f_j, s) \wedge \mathrm{holds}(f, s) \rightarrow \neg \mathrm{holds}(f, \mathrm{result}(a, s))$$

Every sentence of this form tell us about a particular change that an action causes. If we want to assume that actions have as few effects as possible we must have as few of these sentences true as possible. If we have a minimal set of these change sentences then the additional fact that all changes are describable in this way we will generate all the frame axioms we need.

1.4.1 Caveats

Before considering a solution to the frame problem, we first consider some caveats. Kartha and Lifschitz [KL94], and Sandewall [San94] have independently suggested a notion of filtering. Rather than apply a non-monotonic transformation to our initial theory, they propose dividing the theory into two parts. One of these parts is non-monotonically completed, and then the other is conjoined.

We try to avoid this approach. We look at non-monotonic transformations that are applied to the entire theory. We do this for two reasons. Firstly, as an extensive, and well justified theory of non-monotonic reasoning has been developed for this case, and secondly, as there are clear advantages to treating the theory, rather than necessitating a division.

In the rest of this paper we consider deterministic change. That is, an action causes the same effects, if it occurs in two situations where the same fluents hold. We consider the deterministic case as we argue[Cos97] that current approaches to it are lacking. The deterministic setting also seems like a reasonable starting point for any investigation.

1.5 Overview

We first introduce the logical machinery we will use, circumscription and the situation calculus. We then formally define what a solution to the frame problem is semantically. We then show that, with a small modification of circumscription, we can capture this semantic notion.

2 Logical Machinery

Non-monotonic logics are logics where the consequences of a set of sentences are not necessarily consequences of supersets of that set of sentences.

This paper uses circumscription as its non-monotonic machinery. Circumscription is a form of non-monotonic logic introduced by McCarthy [McC80, McC86]. It expresses the non-monotonic consequences of a finitely axiomatizable theory A in a language L under a certain circumscriptive policy as a sentence of second order logic. A circumscriptive policy is a choice of a finite set of formulas of L to minimize[1].

[1] Implicitly we vary all symbols, any particular symbol can be fixed by a suitable set of minimizations following deKleer and Konolige [dKK89].

For a given first order language L, we say an L-structure $\mathfrak{M}$ is preferred to an L-structure $\mathfrak{N}$ if $\mathfrak{M}$ and $\mathfrak{N}$ have the same domain, and the extension of each minimized formula in $\mathfrak{M}$ is a subset of the extension of that formula in $\mathfrak{N}$. We write this pre-order as $\leq_\Phi$ if Φ is the set of minimized formulas. In model theoretic terms, the consequences of the circumscription of a theory A are the sentences true in all minimal models of A under this pre-order "is preferred to".

If a theory A is finitely axiomatizable, then we can express these consequences as a second order sentence[2]. We write α for the conjunction of the sentences in a finite axiomatization of A. We write $\phi_1, \phi_2 \leq \psi_1, \psi_2$, for the formula, $\forall \overline{x}.\phi_1(\overline{x}) \rightarrow \psi_1(\overline{x}) \wedge \forall \overline{y}.\phi_2(\overline{y}) \rightarrow \psi_2(\overline{y})$, where $\overline{x}$ is a tuple of variables equal in number and sort to the free variables of ϕ_1 and ψ_1. Similarly $\overline{y}$ is a tuple of variables equal in number and sort to the free variables of ϕ_2 and ψ_2.

If ϕ_1, ϕ_2 is a pair of formulas, and Z the tuple of all predicate, function and constant symbols in the language L, then we define the circumscription of A minimizing ϕ_1, ϕ_2, written $Circ(A; \phi_1, \phi_2)$ as follows:

$$\alpha \wedge \forall Z_1.[\alpha' \wedge \phi_1', \phi_2' \leq \phi_1, \phi_2] \rightarrow \phi_1, \phi_2 \leq \phi_1', \phi_2'$$

Here $\alpha', \phi_1', \phi_2'$ denote $\alpha[Z/Z_1]$, $\phi_1[Z/Z_1]$, $\phi_2[Z/Z_1]$, the replacement of the symbols Z by variables Z_1, a sequence of predicate, function and object variables agreeing in arity and number with the predicate, function and constant symbols in Z.

2.1 The Language of the Situation Calculus

The situation calculus of order n, L^n_{sit}, is a language of n^{th} order logic with equality, with three non-empty disjoint sorts[3]: fluent names, `fluent`, actions `act` , and situations sit . It contains exactly one predicate constant, holds, in addition to equality, one function constant $\mathrm{result}(a, s)$ which returns a situation, a function $\mathrm{sit}(g)$ from predicates of fluents to situations, and one situation constant S0, plus a set of fluent constants $\overline{f}$, and a set of action constants $\overline{a}$. When we wish to specify the action and fluent constants we write $L^n_{sit}(\overline{f}, \overline{a})$.

No essential difficulties arise if we look at extension of the situation calculus that allows other sorts, and functions from those sorts to fluents and actions. For instance, if we allow a sort *block*, a a block constant A, we can have a fluent, $clear(A)$.

The sorts in the situation calculus will be treated as predicates for the purposes of circumscription.

3 A Natural Solution

We consider a certain subset of the formulas of our language, namely those that imply a change occurred. These we call change formula. We consider the following as our definition of a change formula:

Definition: 1 (change formula)
A change formula is a formula in the language $L^2_{sit}(\overline{f}, \overline{a})$ that is an instance of the following schemas:

$$\forall s. \bigwedge_{f_i \in P} \mathrm{holds}(f_i, s) \wedge \bigwedge_{f_j \in N} \neg\mathrm{holds}(f_j, s) \wedge \mathrm{holds}(f, s) \rightarrow \neg\mathrm{holds}(f, \mathrm{result}(a, s)),$$

[2] We look at the case of minimizing two formulas, the n-ary case is similar

[3] We use s with primes and indices to range over situations, and a and f to range over actions and fluent names, and g to range over unary predicates of fluents.

$$\forall s. \bigwedge_{f_i \in P} \text{holds}(f_i, s) \wedge \bigwedge_{f_j \in N} \neg\text{holds}(f_j, s) \wedge \neg\text{holds}(f, s) \rightarrow \text{holds}(f, \text{result}(a, s)),$$

where f is an element of $\overline{f}$, P and N are subsets of $\overline{f}$, and a is an element of $\overline{a}$.

We note that the axioms that described change, which we considered earlier, had this form. Further, we note that the fluents, and actions in these formulas may have free variables. Therefore, the formula,

$$\begin{array}{c}\forall s.\text{holds}(\text{movable}(x), s) \wedge \neg\text{holds}(\text{at}(x, l), s) \rightarrow \\ \text{holds}(\text{at}(x, l), \text{result}(\text{move}(x, l), s))\end{array}$$

is a change formula.

We now define ab as syntactic sugar, so that conjunctions of ab's correspond to instances of change formulas. More correctly, every change formula is equivalent to a conjunction of ab's.

The change formulas defined above contain two disjoint sets of fluents, P and N. Rather than deal with two sets, we can consider just the case where P is the complement of N. We then only need to consider one set. We use a second order predicate variable g to capture this set. We lose no expressive power by this, as every change formula is equivalent to a (possibly infinite)[4] conjunction of these formulas. We do however gain simplicity by eliminating one of the sets of fluents, which allows for simpler reasoning.

The shorthand $ab(a, f, g)$ can be read as saying that the action a causes the fluent f to change in any situation where the fluents that hold are exactly g. Thus, the set of fluents g corresponds to a set of preconditions.

$$\begin{array}{l} ab(a, f, g) \stackrel{def}{\equiv} \\ \qquad \forall s.(\forall f'.g(f') \equiv \text{holds}(f', s)) \rightarrow (\text{holds}(f, s) \equiv \neg\text{holds}(f, \text{result}(a, s)) \end{array} \tag{1}$$

Now that we have written our change formula as instantiations of a second order logic sentence we can minimize all change formulas by minimizing this sentence.

We can now write our Common Sense Law of Inertia as a second order sentence in terms of the shorthand ab,

$$\begin{array}{l} \forall a\ f\ s.\text{holds}(f, \text{result}(a, s)) \equiv \\ \qquad (\forall g.\neg[ab(a, f, g) \wedge (\forall f'.g(f') \equiv \text{holds}(f', s)] \wedge \text{holds}(f, s)) \vee \\ \qquad (\exists g'.ab(a, f, g') \wedge \forall f'.g'(f') \equiv \text{holds}(f', s))). \end{array}$$

We have written this in the form of explanation closure. That is, it states that if a fluent holds after an action, then either there was no action that cancelled it, whose preconditions were met, or there was an action that caused it, whose preconditions were met.

In this form it is obvious what the role of ab is. It encodes the preconditions for an action causing a change. $ab(a, f, g)$ encodes the preconditions g for an action a to change the fluent f.

The sentence above can be greatly simplified however. Its content is that change is deterministic. Thus we can simplify it to,

$$\begin{array}{l} \forall a\ f_1\ s\ s'.\ (\forall f.\text{holds}(f, s) \equiv \text{holds}(f, s')) \wedge \text{holds}(f_1, \text{result}(a, s)) \\ \qquad \rightarrow \text{holds}(f_1, \text{result}(a, s')). \end{array} \tag{2}$$

[4]The fact that these conjunctions might need to be infinite is not of concern, as we are going to minimize the ab's not use them to represent particular change formulas.

To carry out this simplification we expand the definitions of ab, the shorthand we introduced. This says that if situations agree on all fluents then they behave in the same way.

We now state in model-theoretic terms what the frame assumption is. It is the assumption that there are as few change formulas true as possible.

Definition: 2
The ordering on models of our Common Sense Law of Inertia that captures the assumption of the frame problem is defined as follows. We prefer models where actions have fewer effects. We can measure the effects of an action by the set of formulas of the forms:

$$\begin{array}{l}\forall s. \bigwedge_{f' \in P} \text{holds}(f', s) \wedge \bigwedge_{f'' \in N} \neg\text{holds}(f'', s) \rightarrow \\ \quad (\text{holds}(f, s) \rightarrow \neg\text{holds}(f, \text{result}(a, s)))\end{array}$$

$$\begin{array}{l}\forall s. \bigwedge_{f' \in P} \text{holds}(f', s) \wedge \bigwedge_{f'' \in N} \neg\text{holds}(f'', s) \rightarrow \\ \quad (\neg\text{holds}(f, s) \rightarrow \text{holds}(f, \text{result}(a, s)))\end{array}$$

Let ϕ range over formulas of these forms. Let $\mathfrak{A}, \mathfrak{B}$ be two models of L_{sit}. We say $\mathfrak{A} \leq_{frame} \mathfrak{B}$, if

$$\mathfrak{A} \models \phi \text{ implies } \mathfrak{B} \models \phi$$

We write $A \models_{frame} \psi$ if ψ is true in the minimal models of A under the pre-order $\leq_{frame}$

We note that these models are assumed to obey the rule of explanation closure/Common Sense Law of Inertia/determinism above.

3.1 Circumscription Formalization

We now capture the semantic criterion we considered earlier in circumscription. It is worth noting that our semantic criterion was a preference on models, and thus is already quite close to circumscription. The only difference that we will need to take care of is that circumscription's preference relation on models does not compare models that have different domains, while our semantic preference does. We will overcome this by varying our sorts, and stating that extra objects, not in any of our three sorts, exist.

We now add a (second order) sentence that states that there is an infinite set of objects that are not fluents, situations, or actions. These extra objects exist to fix a technical restriction on circumscription. Circumscriptive consequence corresponds to minimal entailment over pre-orders that do not compare structures whose domains differ. As we will want to compare structures with differing set of situations/fluents/actions, we add a new set of objects, and vary what we consider fluents, situations, and actions to be.

The sentence below, which we call *inf*, states that there is an infinite number of objects that are not actions, situations, or fluents, by encoding first two Peano axioms.

$$\begin{array}{l}\exists f, x. \forall y. f(y) \neq x \wedge \forall y, z. y \neq z \rightarrow f(y) \neq f(z) \wedge \\ \forall y. \neg\texttt{sit}(f(y)) \wedge \forall y. \neg\texttt{fluent}(f(y)) \wedge \forall y. \neg\texttt{act}(f(y)) \wedge \\ \neg\texttt{sit}(x) \wedge \neg\texttt{fluent}(x) \wedge \neg\texttt{act}(x)\end{array} \tag{3}$$

We also now add that there are at least as many objects that are not actions, situations, or fluents, as sets of fluents. We can do this most easily by stating that there is a surjective function from sets of fluents to objects that are neither fluents, actions nor situations.

$$\begin{array}{l} \forall g, g'.(\exists f.g(f) \not\equiv g'(f)) \rightarrow sur(g) \neq sur(g') \\ \forall g.\neg \mathtt{sit}(sur(g)) \wedge \neg \mathtt{fluent}(sur(g)) \wedge \neg \mathtt{act}(sur(g)) \end{array} \tag{4}$$

Definition: 3
$\models_{frame}$ The consequences of a finitely axiomatizable problem theory A in the language L_{sit}, under the frame assumption, written $\models_{frame}$, are the consequences of,

$$Circ\Big(\ A \wedge 2 \wedge 3 \wedge 4; ab\ \Big).$$

It is notable that we vary the sorts `sit`, `fluent` and `act`.

Theorem: 1
The function induced by $\models_{frame}$ solves the frame problem, that is, it agrees with minimal entailment under $\leq_{frame}$, for first order sentences.

Before we prove this theorem, we consider the following lemma.

Lemma: 1
If A is a first order theory in L_{sit}, and $\mathfrak{M}$, $\mathfrak{N}$ models of A such that $\mathfrak{M} \leq_{frame} \mathfrak{N}$, then there is a model $\mathfrak{L}$ such that $\mathfrak{L} \leq_{frame} \mathfrak{N}$, and $\mathfrak{M}$ is elementarily equivalent to $\mathfrak{L}$, where the domain of $\mathfrak{L}$ is contained in $N \cup X$, where X has cardinality $max(2^{\|F\|}, \aleph_0)$, if $\|F\|$ is the cardinality of the sort `fluent`.

This is proved in a later section. Before we prove this we give some intuition.

We now prove the theorem.

We consider the ordering on models $\leq_{frame}$. The ordering on models that our circumscription policy gives is clearly a weaker order. This means that there might be more minimal models in the circumscription order than in the $\leq_{frame}$ order. We show however that for every minimal model in the circumscription order, there is an elementarily equivalent minimal model in the $\leq_{frame}$ order, using the above lemma 1. Suppose there was not. Then there would be models $\mathfrak{A}$ and $\mathfrak{B}$ such that $\mathfrak{A} \leq_{frame} \mathfrak{B}$, but there would be no model $\mathfrak{C}$ that was less in the circumscription order. However, the above lemma implies that such a model exists, a contradiction, which proves our theorem.∎

This implies that for every minimal model under the ordering $\leq_{frame}$ there is an elementarily equivalent minimal model whose domain contains at most $max(2^{\|F\|}, \aleph_0)$ elements . The standard model theory of circumscription tells us that our circumscription finds the minimal models

3.2 The role of sur and *inf*

We now explain the role of varying the sorts, and adding extra objects. This latter change is to ensure that there were enough situations in our models.

The reason that we needed a certain number of situations in every model has to do with a technical deficiency of circumscription. Circumscription, when it is working correctly[5] never changes what domains are possible. If the theory A as models with a certain domain, then the result of circumscribing A will have models with that domain, again modulo minimal models existing. The

[5]That is, when enough minimal models exist

reason for this is that circumscription corresponds to choosing the minimal models of some pre-order in models. These pre-orders on models only compare structures that have the same domain, thus a structure with more situation is never compared to a structure with fewer situations.

We now consider some definitions which make precise what we mean by enough minimal models existing.

Definition: 4
We say that a circumscription policy Φ for L is smooth, with respect to a theory A of L, if every model of A has a minimal model below it, in the ordering $\leq_\Phi$.

We can now formally state that circumscription does not effect what domains are considered possible.

Lemma: 2
If A is a finitely axiomatized theory in L, and A has a model with domain N then, if Φ is a smooth circumscription policy for A, then $Circ(A;\Phi)$ has a model of with domain N.

Proof: We are given a model $\mathfrak{A}$ of A with domain N, and we need to show that there is a minimal model with this domain. We are given that there is a minimal model $\mathfrak{B}$ below $\mathfrak{A}$. If two models are ordered by a circumscription policy, then they have the same domain. Thus the minimal model $\mathfrak{B}$ has domain N, as required. ∎

We now show that for a particular theory A, all the intended models of A have more situations that a particular model of A. This shows that no smooth circumscription policy can correctly pick out the intended models.

Proposition: 1
The theory T, in the language $L_{sit}(\{A\},\{F\})$, is axiomatized as follows,

$$\forall s.\text{holds}(F,s) \rightarrow \text{holds}(F,\text{result}(A,s)).$$

Its intended models are axiomatized as

$$\forall s.\text{holds}(F,s) \equiv \text{holds}(F,\text{result}(A,s)).$$

$$\neg\forall s.\neg\text{holds}(F,s) \rightarrow \text{holds}(F,\text{result}(A,s)).$$

$$\neg\forall s.\text{holds}(F,s) \rightarrow \neg\text{holds}(F,\text{result}(A,s)).$$

That is, exactly two of the four possible effect axioms are true. The two that represent no change. The other two effect axioms are false.

Lemma: 3
The theory T has model whose set of situations has cardinality 1. No intended models of T have a set of situations of cardinality 1.

Proof: Consider the structure that has as its set of situations $\{S_0\}$. Let the assignment to holds be $\{\langle F,S_0\rangle\}$. Let the assignment to result be the identity, and let the constants be interpreted as themselves. This structure is a model of the theory A above.

Consider any intended model. It satisfies the sentences,

$$\neg\forall s.\neg\text{holds}(F,s) \rightarrow \text{holds}(F,\text{result}(A,s)),$$

$$\neg\forall s.\text{holds}(F,s) \rightarrow \neg\text{holds}(F,\text{result}(A,s)).$$

These two sentences imply,

$$\exists s.\neg \mathrm{holds}(F,s) \wedge \neg \mathrm{holds}(F, \mathrm{result}(A,s)),$$

$$\exists s.\mathrm{holds}(F,s) \wedge \mathrm{holds}(F, \mathrm{result}(A,s)),$$

These two sentences imply,

$$\exists s, s'.s \neq s'$$

using the contrapositive of the axiom of extensionality for equality,

$$\forall x, y.x = y \rightarrow \phi(x) \equiv \phi(y),$$

with $\lambda x.\mathrm{holds}(F,x)$ for ϕ.

Thus every intended model has at least two distinct elements. ∎

Thus we have shown that we shall sometimes need to prefer models with more situations in order to get the intended models. To achieve this we can modify circumscription. This approach is taken in detail [Cos97]. We now show that a countable set is sufficient to preserve all first order consequences, if our circumscription policy is a set of first order sentences.

Lemma: 4
Let L be a first order language, and $\mathfrak{A}$,$\mathfrak{B}$ and $\mathfrak{C}$ L-structures. Let Φ be a finite sets of formulas in L, a circumscription policy.

If $\mathfrak{A} \leq_{\Phi} \mathfrak{B}$ then there is a model $\mathfrak{C}$ elementarily equivalent to $\mathfrak{A}$, whose domain is contained in $B \cup X$, where X is a countable new set of objects.

Proof:

We denote the domain of an L-structure $\mathfrak{A}$ as A. We consider the objects $B \cap A$. We construct a model $\mathfrak{C}$, such that $C \subseteq (B \cap A) \cup X$, that agrees with $\mathfrak{A}$ over the objects $B \cap A$, and is elementarily equivalent to $\mathfrak{A}$. This exists, as every substructure can be extended to an elementarily embedding by adding a countable number of objects.

We now notice that this model $\mathfrak{C}$ is $\leq_{\Phi}$ than $\mathfrak{B}$, as $\mathfrak{A} \leq_{\Phi} \mathfrak{B}$, and $\mathfrak{C}$ agrees with $\mathfrak{A}$ on all objects in the intersection of $\mathfrak{A}$ and $\mathfrak{B}$, and the fact that the $\leq_{\Phi}$ relation is defined in terms of subset inclusion on the objects common to both models. ∎

However, we cannot immediately use this lemma, as we are minimizing a second order formula with free second order variables. To get an analogue of this theorem for circumscription policies with second order free variables we would need to add a much larger amount of new objects. However, the pathological cases are rare. We saw that in the previous example every intended model needed at least $2^{||F||}$ situations where $||F||$ is the cardinality of the set of fluents. In general $2^{||F||}$ extra objects is sufficient if the theory A is a first order theory. This is not the case for second order theories, which might contain a sentence that states if there are more than n situations, there are at least m situations where m was a large cardinal.

We first show that for the circumscription policy we consider $2^{||F||}$ suffices for first order theories.

Proposition: 2
If A is a first order theory in L_{sit}, and $\mathfrak{M}$, $\mathfrak{N}$ models of A such that $\mathfrak{M} \leq_{frame} \mathfrak{N}$, then there is a model $\mathfrak{L}$ such that $\mathfrak{L} \leq_{frame} \mathfrak{N}$, and $\mathfrak{M} \equiv \mathfrak{L}$, where the domain of $\mathfrak{L}$ is contained in $N \cup X$, where X has cardinality $max(2^{||F||}, \aleph_0)$.

Proof:

If $\mathfrak{M} \leq_{frame} \mathfrak{N}$ this is because there are fewer change formula true in the structure $\mathfrak{M}$. By the definition of $\leq_{frame}$, the ordering is defined on the objects, and the sets of objects in both models. We add to $\mathfrak{M}$ $2^{||F||}$ new fluent constants, g_n, for every set of fluents n, and write sentences

$$\forall s.\text{holds}(g_n, s) \rightarrow \text{holds}(F, s)$$

for every $F \in n$.

Let X be a set of cardinality $max(2^{||F||}, \aleph_0)$. We construct a model $\mathfrak{L}$, such that $L \subseteq (M \cap N) \cup X$, that agrees with $\mathfrak{A}$ over the objects $M \cap N$, and is elementarily equivalent, in the larger language, to $\mathfrak{M}$. This exists, as every substructure can be extended to an elementarily embedding by adding a the maximum of a countable number of objects, or the cardinality of the underlying language. The cardinality of the language is bounded by $max(2^{||F||}, \aleph_0)$, as we have added $2^{||F||}$ new constants to a countable language. $\square$

4 Causality and Possible Worlds

Causality is sometimes characterized by stating that an action causes an effect if the effect necessarily occurs after the action.

It is natural to interpret this as that an action A causes a change in a fluent F, if it does so in every possible situation.

If we follow this notion it becomes clear that if we are to use effect axioms to represent causality, then our domain of situations must contain all possible situations, rather than all *actual* situations.

Reasoning similar to the above is what justifies Baker's *existence of situations* axioms which we consider in a longer version. These axioms state that if it is consistent that a situation, where a given set of fluents hold, exists, then it does exist.

This view of the role of situations is strongly antithetical with the postulate of induction over situations. This states that the only situations that exist are those that result from a finite sequence of actions beginning at the initial situation.

5 Other Work

The first set of solutions to the frame problem, which proposed preferring change that happened as late in time as possible, were by Lifschitz [Lif86], Shoham [Sho86a, Sho86b, Sho87b, Sho87a, Sho88b, Sho88a, SM88, Sho87c], and Kautz [Kau86, AK85]. These solutions are now grouped under the heading of chronological minimization, a term suggested by Shoham. It is worthwhile noting that these solutions did not change the basic default, the preference was still to assume that change happened as few times as possible, but a preference between defaults at different times was expressed. These solutions all suffered from the flaw that they only worked correctly when predicting the future—temporal projection—the standard example of when they fail is when we are told some fact about the future that we would not have predicted. This was pointed out early on by Lifschitz and Ginsberg.

The next set of solutions was those based on "causal" minimization. Haugh and Lifschitz introduced formalisms with a "causes" predicate that captured the notion that "in the intended model, *all changes in the values of fluents are caused by actions*, whereas in other models this is not the case" [Lif87]. The main drawbacks of this approach are that it does not allow us to write our axioms in the unrestricted situation calculus, rather we must use the "causes" predicate when it is appropriate. Because of this restriction it is not clear how to add domain constraints or ramifications to this formalism. These points were expressed by Baker [Bak91]. The major advance of these theories from the present author's point of view is that they rejected minimizing the total amount of change, and rather minimized causality.

The next important solution was proposed by Baker [Bak91]. His method is more robust than many other methods—it deals correctly with reasoning backwards in time and reasoning about domain constraints. He suggested returning to the idea that the total amount of change is minimized. Baker claims that because the the abnormality predicate takes a situational argument, the meanings of situations should be held constant across models that are compared. He claims that varying the function *result* that picks out the situation after an action is performed, rather than varying what fluents hold at each situation, captures this intuition.

In order for his method to work it is necessary to add an existence of situations axiom, that states that every possible combination of properties holds at some point in time. Baker does not justify this from intuition, but rather states that it is needed to make the formalism work. We show in a longer version that this axiom comes about naturally as a result of minimizing change axioms.

Lin and Shoham [Lin91, LS92] suggested a solution that again returned to minimizing the total amount of change, but they suggested that this should be minimized one situation at a time. Their solution can be explained as follows. For every situation, and possible set of fluents that could hold at that situation, they calculate what the minimal amount of change between that situation and its successors could be. They then add the conditional that if those fluents held at the situation, then the successors will be exactly those that have minimal differences. This approach was extended by Lin and Reiter [LR94]. The major drawback with this approach is that it can become inconsistent when we are given disjunctive information. However, Lin and Shoham show that it behaves correctly for a limited class of theories, essentially those where what holds in the initial situation uniquely determines what is holds in all later situations.

Kartha and Lifschitz [KL95] based on the work of Winslett [Win88], also suggested that a solution could be achieved by considering change one situation at a time. The disadvantages to this work are that it involves changing the language of situation calculus quite drastically. It also is based on circumscribing parts of the theory, and thus treats the theory as if it had extra internal structure. No semantics are given for this extra internal structure. This proposal cannot deal with disjunctions of effect axioms and observations.

In addition to the work that has used non-monotonic logics, Gelfond and Lifschitz have suggested using logic programming, and have introduced various languages $\mathcal{A}$ etc., that have a computational interpretation. Lifschitz [Lif93], and Kartha and Lifschitz [KL94], and Giunchiglia et al. [GKL95] have extended these languages moving away from the computational aspects towards a minimal model paradigm. They give no semantics to their statements about change, and thus it is difficult to evaluate what their theory says about change. Gelfond and Baral [BGPar] have extended $\mathcal{A}$ in the direction of logic programming, while McCain and Turner [MT95] have looked more closely at the declarative interpretation of the sentences of $\mathcal{A}$ and at how domain constraints can imply preconditions, an issue we have not addressed.

Monotonic descriptions of the frame axioms have also been developed by Reiter [Rei91], Pednault [Ped89], Schubert [Sch90], and originally Hass [Has86]. These methods of generating the frame axioms worked on the syntactic form of the theory. As we mentioned earlier they cannot handle ramifications, and thus cannot handle effect axioms that are incompatible

Shanahan's book, [Sha97] and the articles in [GHR94] give a good introduction to the frame problem.

6 Caveats

The methodology this paper proposes for reasoning about change is by no means ideal. In particular it does not address the problems that arise when we reason about non-determinism, concurrent action, strategies, or the qualification problem.

7 Conclusion

We present a simple solution to the frame problem. Our solution can be defined using a slight extension of circumscription. Like McCarthy's original proposal it varies holds, but rather than minimize the amount of change, it minimizes the amount of change formulas.

It applies a single circumscription policy to the entire theory. In this it is superior to the suggestions of Baker, and Kartha and Lifschitz. It correctly handles disjunctions, differing from all previous approaches on certain examples[Cos97].

References

[AK85] J. F. Allen and H. A. Kautz. A model of naive temporal reasoning. In Jerry R. Hobbes and Robert C. Moore, editors, *Formal Theories of the Commonsense World*, pages 251–268. Ablex Publishing Corporation, 1985.

[Bak91] A. B. Baker. Nonmonotonic reasoning in the framework of situation calculus. *Artificial Intelligence*, 49:5–23, 1991.

[BGPar] C. Baral, M. Gelfond, and A. Provetti. Reasoning about actions: Laws, Observations and Hypotheses. *Journal of Logic Programming*, to appear.

[Cos97] Tom Costello. *Non-monotonicity and Change*. PhD thesis, Stanford University, 1997.

[dKK89] J. de Kleer and K. Konolige. Eliminating the Fixed Predicates from a Circumscription. *Artificial Intelligence*, 39:391–398, 1989.

[GHR94] D. M. Gabbay, C. J. Hogger, and J. A. Robinson. *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol IV*. Oxford University Press, 1994.

[Gin87] M. L. Ginsberg. *Readings in Non-monotonic Reasoning*. Morgan Kaufmann, 1987.

[GKL95] E. Giunchiglia, G. N. Kartha, and V. Lifschitz. Actions with Indirect Effects. Working notes of AAAI spring symposium series on extending theories of action., 1995.

[Has86] A. R. Hass. The Case for Domain Specific Frame Axioms. In Frank M. Brown, editor, *The Frame Problem in Artificial Inteligence*, pages 343–348. Morgan Kaufmann, 1986.

[HM86] S. Hanks and D. McDermott. Default reasoning, nonmonotonic logics and the frame problem. In *Proc. National Conference on Artificial Intelligence (AAAI '86)*, pages 328–333, Philadelphia, 1986.

[Kau86] H. Kautz. A logic of persistence. In *Proc. National Conference on Artificial Intelligence (AAAI '86)*, pages 401–405, 1986.

[KL94] G. N. Kartha and V. Lifschitz. Actions with indirect effects. In J. Doyle, E. Sandewall, and P. Torasso, editors, *Principles of Knowledge Representation and Reasoning: Proc. Fourth International Conference (KR '94)*. Morgan Kaufmann, San Francisco, CA, 1994.

[KL95] G. N. Kartha and V. Lifschitz. A simple formalization of actions using circumscription. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '95)*, 1995.

[Lif86] V. Lifschitz. Pointwise circumscription: a preliminary report. In *Proc. National Conference on Artificial Intelligence (AAAI '86)*, pages 406–410, Philadelphia, 1986.

[Lif87] V. Lifschitz. Formal Theories of Action. In F. Brown, editor, *The Frame Problem in Artificial Intelligence*, pages 121–127. Morgan Kaufmann, 1987.

[Lif93] V. Lifschitz. A Language for Describing Actions. In *Working Papers of Commonsense '93*, pages 103–113, 1993.

[Lin91] F. Lin. *A Study of Nonmonotonic Reasoning.* PhD thesis, Stanford University, 1991.

[LR94] F. Lin and R. Reiter. State Constraints Revisited. *Journal of Logic and Computation*, 4(5):655–678, 1994.

[LS92] F. Lin and Y. Shoham. Concurrent actions in the situation calculus. In *Proc. National Conference on Artificial Intelligence (AAAI '92)*, pages 590–595, 1992.

[McC80] J. McCarthy. Circumscription - a Form of Non-Monotonic Reasoning. *Artificial Intelligence*, 13(1 and 2):27–39, 1980.

[McC86] J. McCarthy. Applications of Circumscription to Formalizing Commonsense Knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[MH69] J. McCarthy and P. Hayes. Some Philosophical Problems From the Standpoint of Artificial Intelligence. In D. Michie, editor, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh, UK, 1969.

[MT95] N. McCain and H. Turner. A causal theory of ramifications and qualifications. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '95)*, pages 1978–1984, 1995.

[Ped89] E. Pednault. ADL: Exploring the middle ground between STRIPS and the situation calculus. In *Proc. First International Conference on Principles of Knowledge Representation and Reasoning (KR '89)*, pages 324–332, 1989.

[Rei91] 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.

[San94] E. Sandewall. *Features and Fluents.* Oxford University Press, 1994.

[Sch90] L. Schubert. Monotonic solution of the frame problem in the situation calculus: An efficient method for worlds with fully specified actions. In H. E. Kyburg, Ronald P. Loui, and Greg N. Carlson, editors, *Knowledge Representation and Defeasible Reasoning*, volume Volume 5 of *Studies in Cognitive Systems*, pages 23–67. Kluwer Academic Publishers, Dordrecht / Boston / London, 1990.

[Sha97] Murray Shanahan. *Solving the Frame Problem, a mathematical investigation of the common sense law of inertia.* M.I.T. Press, 1997.

[Sho86a] Y. Shoham. Chronological Ignorance: Time, Nonmonotonicity, Necessity and Causal Theories. In *Proc. National Conference on Artificial Intelligence (AAAI '86)*, pages 389–393, 1986.

[Sho86b] Y. Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence.* PhD thesis, Yale University, 1986.

[Sho87a] Y. Shoham. Nonmonotonic logics: meaning and utility. In *Proc. Tenth International Joint Conference on Artificial Intelligence (IJCAI '87)*, San Mateo, CA., 1987. Morgan Kaufmann.

[Sho87b] Y. Shoham. A semantical approach to nonmonotonic logics. In *Proc. 2nd IEEE Symp. on Logic in Computer Science*, pages 275–279, 1987. Reprinted in [Gin87].

[Sho87c] Y. Shoham. Temporal Logics in AI: Semantical and Ontological Considerations. *Artificial Intelligence*, 33(1):89–104, September 1987.

[Sho88a] Y. Shoham. Chronological Ignorance: Experiments in Nonmonotonic Temporal Reasoning. *Artificial Intelligence*, 36:271–331, 1988.

[Sho88b] Y. Shoham. *Reasoning about Change.* MIT Press, Cambridge, USA, 1988.

[SM88] Y. Shoham and D. McDermott. Problems in Formal Temporal Reasoning. *Artificial Intelligence*, 36(1):49–62, August 1988.

[Win88] M. Winslett. Reasoning about action using a possible models approach. In *Proc. National Conference on Artificial Intelligence (AAAI '88)*, pages 89–93, 1988.