Toward a Metatheory of Action

Vladimir Lifschitz
Department of Computer Sciences
and Department of Philosophy
University of Texas at Austin
Austin, TX 78712

Abstract

We present a formalization of the blocks world on the basis of the situation calculus and circumscription, and investigate its mathematical properties. The main theorem describes the effect of the circumscription which solves the frame problem in the presence of ramifications. The theorem is quite general, in the sense that it is applicable not only to the blocks world, but to a whole class of domains involving situations and actions. Its statement does not mention anything specific for the domain of blocks. Instead, it lists assumptions about purely formal, mostly syntactic, properties of the axiom set.

1 INTRODUCTION

In this paper we present a formalization of the blocks world on the basis of the situation calculus and circumscription, and investigate its mathematical properties.

The formalization is not particularly original; it is based on the approach of [Baker, 1989][1], and is in some ways similar to the formulation from Section 6 of [Baker and Ginsberg, 1989]. One difference is that we use a more abstract formalization of states, along the lines of [Lifschitz, 1990], which allows us, for instance, to do without any specific assumptions about the number of available blocks.

The main novelty here is a theorem, which describes the effect of the circumscription used for solving the frame problem in the presence of ramifications, and confirms the adequacy of this solution. The theorem is quite general, in the sense that it is applicable not only to the blocks world, but to a whole class of domains involving situations and actions. Its statement does not mention anything specific for the domain of blocks. Instead, it lists assumptions about purely formal, mostly syntactic, properties of the axiom set. One of the assumptions, for instance, is that the axiom set includes a certain form of the commonsense law of inertia.

Such "metatheoretical" investigation of commonsense knowledge gives us not merely one successful formalization of one particular domain, but a class of successful formalizations. This approach has two advantages.

First, any new commonsense domain that we want to formalize may happen to have a formalization which belongs to a class already familiar to us. Then the confirmation of the adequacy of this formalization will come from a metatheorem proved earlier.

Second, databases of common sense, like any other databases, will need to be updated, and it is crucial that an update be allowed only when the designer of the database has a clear understanding of the effect of updates of that type. Hopefully, this can be achieved by implementing the requirement that the updated database belong to a class of theories whose properties are well understood. The metalevel investigation of formalizations of common sense will be instrumental for solving this problem.

The main theorem, stated in Section 3 of this paper and proved in Section 4, shows how these ideas apply to circumscriptive theories of action. Prior to discussing this theorem, we describe in Section 2 a formalization of the blocks world which satisfies its

[1] This was the first successful application of circumscription [McCarthy, 1986] to the frame problem for actions with indirect effects ("ramifications").

conditions and thus plays the role of a motivating example. The version of the blocks world used in this example is extremely simple; the only available action is moving one block on top of another. But the theorem covers more complex domains as well. We discuss its possibilities and limitations in Sections 3.3 and 3.4; we show, in particular, that it is applicable to the blocks world in which blocks have colors and can be not only moved, but also painted.

2 THE BLOCKS WORLD EXAMPLE

2.1 CIRCUMSCRIPTIVE THEORIES

The blocks world will be formalized as a *circumscriptive theory*, in the following sense of this term. Consider a language of classical predicate calculus with equality—one-sorted or many-sorted, first-order or higher-order. A circumscriptive theory consists of a finite set Γ of sentences of this language, called the *axioms* of the theory, and a finite nonempty set Δ of expressions of the form

$$\textbf{circ}\ P\ \textbf{var}\ Z_1, \ldots, Z_n, \tag{1}$$

where P is a predicate constant, and $Z_1, \ldots, Z_n$ are predicate or function constants. These expressions will be called the *policy declarations* of the theory (because they determine its "circumscription policy"). A circumscriptive theory (Γ, Δ) represents the classical axiomatic theory whose axioms are the sentences[2]

$$\mathrm{CIRC}\left[\bigwedge_{A \in \Gamma} A; P; Z_1, \ldots, Z_n\right] \tag{2}$$

for all declarations (1) from Δ. In particular, by a *model* of (Γ, Δ) we mean a model of the formulas (2). A *theorem* of (Γ, Δ) is a sentence which is true in all its models.

The language of a circumscriptive theory can be many-sorted and higher order. In most applications of circumscription, higher order variables do not occur in the axioms, and are only needed for forming the circumscription formula. Baker ([Baker, 1989], Section 5) noticed, however, that predicate variables are useful for stating his "existence of situations" axioms, and we follow this approach in Section 2.2.

Extending the definition of circumscription to the case when some or all arguments of the circumscribed predicate are higher order variables is straightforward.

[2] CIRC$[A; P; Z]$ stands for the result of circumscribing the predicate P in the sentence A with Z allowed to vary [McCarthy, 1986], [Lifschitz, 1985], [Genesereth and Nilsson, 1987].

2.2 AXIOMS FOR SITUATIONS AND FLUENTS

First we describe the part of the axiom set which deals with situations and fluents; axioms for actions will be added in Section 2.3. This part of the formalization uses variables of three sorts: for blocks $(x, y, z, x_1, y_1, z_1, \ldots)$, for situations $(s, s_1, \ldots)$, and for propositional fluents $(f, f_1, \ldots)$. The formula $Holds(f, s)$ expresses that the value of f in the situation s is *true*. The specific fluents that are of interest to us in this example will be represented using the function constants *On* and *Ontable*. The terms $On(x, y)$ and $Ontable(x)$ represent distinct fluents for different values of x and y:

$$\begin{gathered} On(x_1, y_1) = On(x_2, y_2) \supset x_1 = x_2 \wedge y_1 = y_2, \\ Ontable(x) = Ontable(y) \supset x = y, \\ On(x, y) \neq Ontable(z). \end{gathered} \tag{3}$$

The fluents $On(x, y)$ and $Ontable(x)$ play the role of a "coordinate frame" in the space of situations, in the sense that specific configurations of blocks can be described by combinations of values of these "frame fluents" [McCarthy and Hayes, 1969]. For instance, the configuration in which all blocks are on the table can be characterized by saying that the value of each of the fluents $Ontable(x)$ is *true*.[3]

By a "state" we understand a function assigning values to all frame fluents. When the frame fluents are propositional, a state is a truth-valued function on the set of frame fluents, and we will identify it with the set of fluents to which it assigns the value *true*. Every situation defines a certain state—the set of all frame fluents that are true in this situation. But one state may correspond to many different situations. For instance, two different situations s_1 and s_2 may share the property that all blocks are located on the table; perhaps the difference between s_1 and s_2 is in the exact locations of blocks, or these situations may correspond to different instants of time.

To formalize these ideas, we use the unary predicate *Frame*, which singles out the frame fluents.[4] It is characterized by the postulates:

$$\begin{gathered} Frame(On(x, y)), \\ Frame(Ontable(x)). \end{gathered} \tag{4}$$

We will circumscribe it, to ensure that the frame contains no fluents other than these.

Since states are sets of frame fluents, they can be represented by unary predicates. We will use $\sigma, \sigma_1, \ldots$ as unary predicate variables whose argument is a fluent

[3] See [Lifschitz, 1990] for a detailed discussion of the role of frames and of possible approaches to formalizing this concept.

[4] For a detailed discussion of this approach to formalizing frames, see [Lifschitz, 1990].

variable. We write $IsState(\sigma)$ for

$$\forall f[\sigma(f) \supset Frame(f)],$$

and $State[s]$ for

$$\lambda f[Frame(f) \wedge Holds(f, s)].$$

Obviously, the formula

$$\forall s\ IsState(State[s])$$

is universally valid.

Given a state σ, we can ask whether there exist a situation s such that $State[s] = \sigma$. Many combinations of values of frame fluents are impossible to realize physically, and even difficult to imagine. A block cannot be located in two places at once; a block cannot be located on its own top; it is impossible to build "circular" configurations, when, for instance, x is located on top of y, y on top of z, and z on top of x. However, we do not think of every situation as necessarily realized at some point in the actual course of events; consequently, there would be nothing wrong with admitting "ideal" situations, corresponding to physically impossible states. In fact, postulating such "ideal" situations is useful, because it allows us to guarantee the existence of a variety of different situations without going into the detailed study of what is physically possible and what is not.

The formula

$$IsState(\sigma) \supset \exists s(\sigma = State[s])$$

is a very strong "existence of situations" axiom; it asserts that *every* state corresponds to at least one situation, so that the frame fluents are completely independent. Alternatively, this assumption can be stated in the Skolemized form:

$$IsState(\sigma) \supset \sigma = State[Sit(\sigma)]. \tag{5}$$

The function Sit maps a state into one of the corresponding situations.

In our formalization of the blocks world, we do not want to go quite this far in allowing "ideal" situations. We will assume the constraints on possible combinations of the values of frame fluents, according to which a block cannot be located in two places at once:

$$\begin{aligned} &Holds(On(x, y_1), s) \wedge Holds(On(x, y_2), s) \\ &\qquad\qquad\qquad \supset y_1 = y_2, \\ &Holds(On(x, y), s) \supset \neg Holds(Ontable(x), s). \end{aligned} \tag{6}$$

Then (5) needs to be replaced by the corresponding "default":

$$IsState(\sigma) \wedge \neg Inconsistent(\sigma) \supset \sigma = State[Sit(\sigma)], \tag{7}$$

where *Inconsistent* is the new predicate, which will be circumscribed.

We expect that this circumscription will lead to the conclusion that the constraints (6) represent all existing dependencies between the values of frame fluents. One more assumption is needed in order to make this work:

$$\begin{aligned} &IsState(\sigma_1) \wedge IsState(\sigma_2) \wedge Sit(\sigma_1) = Sit(\sigma_2) \\ &\qquad\qquad\qquad \supset \sigma_1 = \sigma_2. \end{aligned} \tag{8}$$

This axiom guarantees that the cardinality of the universe of situations is sufficiently large.

To sum up, the axioms for situations and fluents are the universal closures of the formulas (3), (4) and (6)–(8).

2.3 AXIOMS FOR ACTIONS

Now we extend the language used in Section 2.2 by variables for actions $a, a_1, \ldots$. We also add the binary function constant *Result*, whose arguments are an action and a situation and whose value is a situation, and the binary function constant *Move*, whose arguments are blocks and whose value is an action. Intuitively, the term $Move(x, y)$ represents the action of placing x on top of y. These terms represent distinct actions for different values of x and y:

$$Move(x_1, y_1) = Move(x_2, y_2) \supset x_1 = x_2 \wedge y_1 = y_2. \tag{9}$$

We will use the atomic formula $Possible(a, s)$ to express that it is possible to carry out the action a in the situation s. The following axiom describes the effect of $Move(x, y)$:

$$\begin{aligned} &Possible(Move(x, y), s) \\ &\quad \supset Holds(On(x, y), Result(Move(x, y), s)). \end{aligned} \tag{10}$$

A sufficient condition for the possibility of $Move(x, y)$ is given by the axiom:

$$\begin{aligned} &\neg\exists z\ Holds(On(z, x), s) \\ &\quad \wedge \neg\exists z\ Holds(On(z, y), s) \\ &\qquad \wedge x \neq y \\ &\qquad\qquad \supset Possible(Move(x, y), s). \end{aligned} \tag{11}$$

Finally, we include the following form of the commonsense law of inertia:

$$\begin{aligned} &Frame(f) \wedge Possible(a, s) \wedge \neg Noninertial(f, a, s) \\ &\qquad \supset [Holds(f, Result(a, s)) \equiv Holds(f, s)]. \end{aligned} \tag{12}$$

Here *Noninertial* is a new predicate constant, which will be circumscribed.

2.4 POLICY DECLARATIONS

By BW we denote the circumscriptive theory whose axioms are the universal closures of the formulas (3), (4) and (6)–(12), and whose policy declarations are:

$$\begin{aligned} &\textbf{circ}\ Frame\ \textbf{var}\ Inconsistent, Noninertial, \\ &\textbf{circ}\ Inconsistent\ \textbf{var}\ Holds, Result, Possible, \\ &\qquad\qquad\qquad\qquad Noninertial, \\ &\textbf{circ}\ Noninertial\ \textbf{var}\ Result. \end{aligned} \tag{13}$$

Notice that *Inconsistent* is varied when *Frame* is circumscribed, but not the other way around; in this sense, *Frame* is minimized at a higher priority than *Inconsistent*. This is because we want to think of the extent of the frame as already fixed when we decide which states are consistent and which are not. Furthermore, this circumscription policy minimizes *Noninertial* at a lower priority than the other two circumscribed predicates, because we think of the structure of the space of situations as already determined when the effects of actions are described. Since the intention of the law of inertia is to characterize the result of executing a, the function *Result* is varied as *Noninertial* is circumscribed.

2.5 EFFECT OF THE CIRCUMSCRIPTIONS

The main theorem, applied to the theory *BW*, will show that the set of theorems of *BW* includes certain complete characterizations (explicit definitions) of all three circumscribed predicates.

The definition of *Frame* is quite simple:

$$\begin{aligned}&\mathit{Frame}(f)\\ &\quad \equiv \exists xy[f = \mathit{On}(x,y)] \lor \exists x[f = \mathit{Ontable}(x)].\end{aligned}$$

This is indeed what we expected to get when we postulated that all fluents of the forms $\mathit{On}(x,y)$ and $\mathit{Ontable}(x)$ belong to the frame, and circumscribed *Frame*. Notice that this equivalence is the result of circumscribing *Frame* relative to the conjunction F of (the universal closures of) the axioms (4), so that it can be written in the form

$$\mathrm{CIRC}[F; \mathit{Frame}]. \tag{14}$$

To describe the effect of circumscribing the predicate *Inconsistent*, we need the following notation: $C_0(\sigma)$ stands for[5]

$$\begin{aligned}&\forall x y_1 y_2[\sigma(\mathit{On}(x,y_1)) \land \sigma(\mathit{On}(x,y_2)) \supset y_1 = y_2]\\ &\quad \land \forall xy[\sigma(\mathit{On}(x,y)) \supset \neg\sigma(\mathit{Ontable}(x))].\end{aligned}$$

This formula expresses the constraint on the state σ similar to the constraint we have imposed on situations: A block cannot be in two places at once. The characterization of *Inconsistent* given by the theory *BW* is:

$$\mathit{Inconsistent}(\sigma) \equiv \mathit{IsState}(\sigma) \land \neg C_0(\sigma). \tag{15}$$

Thus the state is consistent unless it requires some block to be in two places at once.

For determining the minimal extent of *Noninertial*, we need the following notation: $R_0(a,f)$ stands for

$$\exists xy(a = \mathit{Move}(x,y) \land f = \mathit{On}(x,y)).$$

[5]The reason why we chose this particular symbol is that it is convenient in the general framework of the main theorem. The same can be said about the symbol R_0 introduced below.

This formula expresses that a "causes" f. By $\mathit{Affected}(f,a,\sigma)$ ("the fluent f is affected by the action a in the state σ") we denote the formula

$$\begin{aligned}&\forall\sigma_1[\mathit{IsState}(\sigma_1) \land C_0(\sigma_1)\\ &\quad \land \forall f_1(R_0(a,f_1) \supset \sigma_1(f_1))\\ &\qquad \supset \neg(\sigma_1(f) \equiv \sigma(f))].\end{aligned}$$

The antecedent of this conditional expresses that σ_1 is a consistent state that can occur after the execution of a. Thus we say that f is affected by a in the state σ if the value of f in any such state σ_1 is different from its value in σ.

Using this notation, we can express the characterization of *Noninertial* given by *BW* by the formula:

$$\begin{aligned}&\mathit{Noninertial}(f,a,s) \equiv\\ &\mathit{Frame}(f) \land \mathit{Possible}(a,s) \land \mathit{Affected}(f,a,\mathit{State}[s]).\end{aligned} \tag{16}$$

To illustrate the role of this conclusion, we will show that it allows us to prove the "frame axiom"

$$\begin{aligned}&\mathit{Possible}(\mathit{Move}(x_1,y_1),s) \land x_1 \neq x_2\\ &\quad \supset [\mathit{Holds}(\mathit{On}(x_2,y_2), \mathit{Result}(\mathit{Move}(x_1,y_1),s))\\ &\qquad \equiv \mathit{Holds}(\mathit{On}(x_2,y_2),s)].\end{aligned} \tag{17}$$

Define

$$\mathit{About}(f,x) \equiv f = \mathit{Ontable}(x) \lor \exists y(f = \mathit{On}(x,y)).$$

The axioms of *BW* imply

$$\begin{aligned}&\mathit{IsState}(\sigma) \land C_0(\sigma) \land x_1 \neq x_2 \supset\\ &\quad \neg\mathit{Affected}(\mathit{On}(x_2,y_2), \mathit{Move}(x_1,y_1), \sigma),\end{aligned}$$

because, whenever $\mathit{IsState}(\sigma) \land C_0(\sigma) \land x_1 \neq x_2$, we can get a counterexample to

$$\mathit{Affected}(\mathit{On}(x_2,y_2), \mathit{Move}(x_1,y_1), \sigma)$$

by taking

$$\sigma_1 = \lambda f[(\sigma(f) \land \neg\mathit{About}(f,x_1)) \lor f = \mathit{On}(x_1,y_1)].$$

Now, using (6), we conclude:

$$\begin{aligned}&x_1 \neq x_2\\ &\quad \supset \neg\mathit{Affected}(\mathit{On}(x_2,y_2), \mathit{Move}(x_1,y_1), \mathit{State}[s]).\end{aligned}$$

Then, by (16),

$$\begin{aligned}&x_1 \neq x_2\\ &\quad \supset \neg\mathit{Noninertial}(\mathit{On}(x_2,y_2), \mathit{Move}(x_1,y_1), s)),\end{aligned}$$

and the formula (17) follows from the law of inertia (12).

Another "frame axiom,"

$$\begin{aligned}&\mathit{Possible}(\mathit{Move}(x_1,y_1),s) \land x_1 \neq x_2\\ &\quad \supset [\mathit{Holds}(\mathit{Ontable}(x_2), \mathit{Result}(\mathit{Move}(x_1,y_1),s))\\ &\qquad \equiv \mathit{Holds}(\mathit{Ontable}(x_2),s)],\end{aligned}$$

can be proved in a similar way.

3 MAIN THEOREM

Now we want to look at the formalization of the blocks world described above from a more general point of view.

3.1 SETTING THE STAGE FOR THE MAIN THEOREM

We assume a many-sorted language L, containing object variables for situations (s,s_1,...), for propositional fluents (f,f_1,...), for actions (a,a_1,a_2,...), and possibly object variables of other, *domain-dependent* sorts. L may contain higher order variables. It is assumed to contain variables σ,σ_1,σ_2,... for properties of fluents (that is, unary predicates with a fluent argument).

In the blocks world example, there is one domain-dependent sort—blocks.

L is assumed to contain the following function and predicate constants:

- the binary predicate *Holds*, whose arguments are a fluent and a situation,
- the unary predicate *Frame*, whose argument is a fluent,
- the unary predicate *Inconsistent*, whose argument is a property of fluents,
- the binary function *Result*, whose arguments are an action and a situation, and whose value is a situation,
- the binary predicate *Possible*, whose arguments are an action and a situation,
- the ternary predicate *Noninertial*, whose argument are a fluent, an action and a situation.

These constants will be called *essential*. Besides the essential constants, L is assumed to contain the unary function constant *Sit*, whose argument is a property of fluents, and whose value is a situation.

L may also contain other object, function and predicate constants.

In the blocks world example, there are 3 additional constants: the function constants *On*, *Ontable* and *Move*.

Notice that the abbreviations *IsState* and *State*, introduced in Section 2.2, can be used in any language L satisfying these conditions.

By T we denote a circumscriptive theory in the language L. The axiom set of T may contain any sentences without essential constants—we will call these axioms *inessential*—and it is assumed to contain certain *essential* axioms, described below.

In the blocks world example, the inessential axioms are the universal closures of (3) and (9).

There are 7 essential axioms:

1. F ("the axiom for *Frame*"), which is assumed to contain no essential constants other than *Frame*.

In the blocks world example, F is the conjunction of the universal closures of the formulas (4).

2. S ("the axiom for *Sit*") is the universal closure of (7), that is,

$$\begin{aligned}&\forall\sigma.\mathit{IsState}(\sigma)\wedge\neg\mathit{Inconsistent}(\sigma)\\ &\qquad\supset\sigma=\mathit{State}[\mathit{Sit}(\sigma)].\end{aligned}$$

3. U ("the uniqueness axiom for *Sit*") is the universal closure of (8), that is,

$$\begin{aligned}&\forall\sigma_1\sigma_2.\mathit{IsState}(\sigma_1)\wedge\mathit{IsState}(\sigma_2)\\ &\qquad\wedge\mathit{Sit}(\sigma_1)=\mathit{Sit}(\sigma_2)\\ &\qquad\qquad\supset\sigma_1=\sigma_2.\end{aligned}$$

4. C ("the domain constraint") is

$$\forall s.C_0(\mathit{State}[s]),$$

where $C_0(\sigma)$ is a formula containing no essential constants and no free variables other than σ.

In the blocks world example, $C_0(\sigma)$ is selected as in Section 2.5. Let us see what C is for this choice of C_0. The first conjunctive term of C is (the universal closure of)

$$\mathit{State}[s](\mathit{On}(x,y_1))\wedge\mathit{State}[s](\mathit{On}(x,y_2))\supset y_1=y_2,$$

that is,

$$\begin{aligned}&\mathit{Frame}(\mathit{On}(x,y_1))\\ &\quad\wedge\mathit{Holds}(\mathit{On}(x,y_1),s)\\ &\quad\quad\wedge\mathit{Frame}(\mathit{On}(x,y_2))\\ &\quad\quad\quad\wedge\mathit{Holds}(\mathit{On}(x,y_2),s)\\ &\quad\quad\quad\quad\supset y_1=y_2.\end{aligned}$$

In the presence of (4), this is equivalent to the first of the formulas (6). Similarly, the remaining part of C gives the second of these formulas. Consequently, including C in the axiom set is equivalent to including (6).

5. R ("the axiom for *Result*") is

$$\begin{aligned}&\forall afs.\mathit{Possible}(a,s)\wedge R_0(a,f)\\ &\qquad\supset\mathit{Holds}(f,\mathit{Result}(a,s)),\end{aligned}$$

where $R_0(a,f)$ is a formula containing no essential constants and no free variables other than a and f.

In the blocks world example, $R_0(a,f)$ is defined as in Section 2.5. In this case, R is equivalent to

$$\begin{aligned}&\forall afsxy.\mathit{Possible}(a,s)\\ &\quad\wedge a=\mathit{Move}(x,y)\\ &\quad\quad\wedge f=\mathit{On}(x,y)\\ &\quad\quad\quad\supset\mathit{Holds}(f,\mathit{Result}(a,s)),\end{aligned}$$

or

$$\begin{aligned}&\forall sxy.\mathit{Possible}(\mathit{Move}(x,y),s)\\ &\quad\supset\mathit{Holds}(\mathit{On}(x,y),\mathit{Result}(\mathit{Move}(x,y),s)),\end{aligned}$$

which is the universal closure of (10).

6. P ("the axiom for *Possible*") is the formula

$$\forall as.P_0(a, State[s]) \supset Possible(a, s),$$

where $P_0(a, \sigma)$ is a formula containing no essential constants and no free variables other than a and σ.

In the blocks world example, $P_0(a, \sigma)$ is

$$\begin{aligned}&\exists xy.a = Move(x, y)\\ &\quad \wedge\neg\exists z[\sigma(On(z, x))]\\ &\quad\quad \wedge\neg\exists z[\sigma(On(z, y))]\\ &\quad\quad\quad \wedge x \neq y.\end{aligned}$$

Then P is equivalent to

$$\begin{aligned}&\forall asxy.a = Move(x, y)\\ &\quad \wedge\neg\exists z[State[s](On(z, x))]\\ &\quad\quad \wedge\neg\exists z[State[s](On(z, y))]\\ &\quad\quad\quad \wedge x \neq y\\ &\quad\quad\quad\quad\quad \supset Possible(a, s),\end{aligned}$$

or

$$\begin{aligned}&\forall sxy.\neg\exists z[State[s](On(z, x))]\\ &\quad \wedge\neg\exists z[State[s](On(z, y))]\\ &\quad\quad \wedge x \neq y\\ &\quad\quad\quad \supset Possible(Move(x, y), s).\end{aligned}$$

In the presence of (4), this can be rewritten as

$$\begin{aligned}&\forall sxy.\neg\exists z\, Holds(On(z, x), s)\\ &\quad \wedge\neg\exists z\, Holds(On(z, y), s)\\ &\quad\quad \wedge x \neq y\\ &\quad\quad\quad \supset Possible(Move(x, y), s),\end{aligned}$$

which is the universal closure of (11).

7. I ("the commonsense law of inertia") is the universal closure of (12), that is,

$$\begin{aligned}&\forall fas.Frame(f)\\ &\quad \wedge Possible(a, s)\\ &\quad\quad \wedge\neg Noninertial(f, a, s)\\ &\quad\quad\quad \supset [Holds(f, Result(a, s)) \equiv Holds(f, s)].\end{aligned}$$

Finally, the circumscription policy of T is assumed to be (13).

3.2 STATEMENT OF THE MAIN THEOREM

The formulas (14)–(16), used in Section 2.5 for characterizing the extensions of the circumscribed predicates, make sense not only in the language of BW, but in any language L of the kind described in Section 3.1. The main theorem asserts that, under certain assumptions about $C_0(\sigma)$ and $R_0(a, f)$, the circumscriptions represented by the policy (13) have exactly the same effect as adding the formulas (14)–(16) to the axioms of T.

The following two conditions have to be imposed on $C_0(\sigma)$ and $R_0(a, f)$:

Condition A. The inessential axioms of T and the axiom F imply

$$R_0(a, f) \supset Frame(f). \tag{18}$$

Condition B. There exists a formula

$$Compatible(f_1, f_2),$$

containing no essential constants and no free variables other than f_1 and f_2, such that the inessential axioms of T imply:

1. $Compatible(f_1, f_2) \equiv Compatible(f_2, f_1)$.
2. $C_0(\sigma) \equiv$
   $\forall f_1 f_2[\sigma(f_1) \wedge \sigma(f_2) \supset Compatible(f_1, f_2)]$.
3. $R_0(a, f_1) \wedge R_0(a, f_2) \supset Compatible(f_1, f_2)$.

Let us check that BW satisfies these conditions. For this theory, (18) is

$$\exists xy(a = Move(x, y) \wedge f = On(x, y)) \supset Frame(f),$$

which is a consequence of (4). We can take

$$\begin{aligned}&Compatible(f_1, f_2)\\ &\quad \equiv \forall x[About(f_1, x) \wedge About(f_2, x) \supset f_1 = f_2]\end{aligned}$$

(*About* is defined in Section 2.5.) Part 1 of Condition B is obvious. To prove part 2, notice that its right-hand side can be written in the form

$$\begin{aligned}&\forall f_1 f_2 x[\sigma(f_1) \wedge \sigma(f_2)\\ &\quad \wedge About(f_1, x) \wedge About(f_2, x)\\ &\quad\quad\quad \supset f_1 = f_2],\end{aligned}$$

or

$$\begin{aligned}&\forall f_1 f_2 x[\sigma(f_1) \wedge \sigma(f_2)\\ &\quad\quad \wedge f_1 = Ontable(x) \wedge f_2 = Ontable(x)\\ &\quad\quad\quad \supset f_1 = f_2]\\ &\wedge\forall f_1 f_2 xy[\sigma(f_1) \wedge \sigma(f_2)\\ &\quad\quad \wedge f_1 = On(x, y) \wedge f_2 = Ontable(x)\\ &\quad\quad\quad \supset f_1 = f_2]\\ &\wedge\forall f_1 f_2 xy[\sigma(f_1) \wedge \sigma(f_2)\\ &\quad\quad \wedge f_1 = Ontable(x) \wedge f_2 = On(x, y)\\ &\quad\quad\quad \supset f_1 = f_2]\\ &\wedge\forall f_1 f_2 xy_1 y_2[\sigma(f_1) \wedge \sigma(f_2)\\ &\quad\quad \wedge f_1 = On(x, y_1) \wedge f_2 = On(x, y_2)\\ &\quad\quad\quad \supset f_1 = f_2].\end{aligned}$$

The first conjunctive term is trivial. The second is equivalent to

$$\begin{aligned}&\forall xy[\sigma(On(x, y)) \wedge \sigma(Ontable(x))\\ &\quad\quad \supset On(x, y) = Ontable(x)].\end{aligned}$$

In the presence of (3), this is equivalent to

$$\forall xy[\sigma(On(x, y)) \supset \neg\sigma(Ontable(x))],$$

which is one half of $C_0(\sigma)$. The third conjunctive term is equivalent to the second. The fourth term is equivalent to

$$\begin{aligned}&\forall xy_1 y_2[\sigma(On(x, y_1)) \wedge \sigma(On(x, y_2))\\ &\quad\quad \supset On(x, y_1) = On(x, y_2)].\end{aligned}$$

In the presence of (3), this is equivalent to

$$\forall x y_1 y_2 [\sigma(On(x, y_1)) \wedge \sigma(On(x, y_2)) \supset y_1 = y_2],$$

which gives the remaining half of $C_0(\sigma)$. Finally, part 3 of Condition B is

$$\begin{aligned}&\exists xy(a = Move(x, y)\\ &\quad \wedge f_1 = On(x, y))\\ &\quad\quad \wedge \exists xy(a = Move(x, y) \wedge f_2 = On(x, y))\\ &\quad\quad\quad\quad \supset Compatible(f_1, f_2).\end{aligned}$$

In the presence of (9), this is equivalent to the universal closure of

$$\begin{aligned}a = Move(x, y) \wedge f_1 = On(x, y) \wedge f_2 = On(x, y)\\ \supset Compatible(f_1, f_2)),\end{aligned}$$

which immediately follows from the definition of *Compatible*.

In the statement of the theorem, T is a circumscriptive theory of the kind described in Section 3.1.

Theorem. *If T satisfies Conditions A and B, then it is equivalent to the conjunction of its axioms and the formulas (14)–(16).*

3.3 DISCUSSION

We arrived at the class of theories described in Sections 3.1 and 3.2 by generalizing a single example—the theory of the blocks world from Section 2. As a result, all these theories share a number of "family traits," inherited from their ancestor. The blocks world example has no initial situation or initial conditions; accordingly, the assumptions of the main theorem make it impossible to have initial conditions in the axiom set. (Initial conditions contain the predicate *Holds*, so that they cannot be included among the inessential axioms. On the other hand, they do not have any of the 7 forms that the essential axioms are allowed to have.) The immediate effect of $Move(x, y)$ is to make a certain fluent true, rather than false; accordingly, the main theorem assumes that the changes caused by all actions are "positive." The blocks world example does not address the qualification problem; accordingly, the circumscriptive theories covered by the main theorem do not deal with it either. There are other limitations.

It seems, however, that it will not be difficult to prove analogs and extensions of the main theorem that overcome many of these limitations. We know, for instance, that the approach used in the blocks world example can handle initial conditions and temporal projection; it should be possible then to describe initial conditions in an abstract form, as one more kind of "essential axioms," and prove the main theorem for systems with such axioms. It may be possible to prove similar theorems for some formalizations of action that include continuous time and concurrency.[6] Moreover, there can be many kinds of mathematical results confirming the adequacy of formalizations; results about the extents of the circumscribed predicates, as in the main theorem, represent only one of them. Assuming that the initial conditions provide the values of all frame fluents in the initial situation, we may be able to prove, for instance, that T decides every instance of the temporal projection problem. Hopefully, the main theorem can serve as a starting point for developing the theory of action along the lines of this metamathematical approach.

3.4 PAINTING BLOCKS

To illustrate the possibilities of the metamathematical approach, we will apply the main theorem to an extension of the blocks world example in which blocks can be not only moved, but also painted. We will see that, with small amount of additional work, we can use the main theorem to determine the effect of circumscription in the enhanced theory.

The theory BW is extended as follows. Variables for colors $c, c_1, \ldots$ are added to the language, along with the binary predicate *Color*, whose arguments are a block and a color, and the binary function *Paint*, whose arguments are a block and a color also, and whose value is an action. The new axioms are:

$$\begin{gathered}Color(x_1, c_1) = Color(x_2, c_2) \supset x_1 = x_2 \wedge c_1 = c_2,\\ Color(x, c) \neq On(y, z),\\ Color(x, c) \neq Ontable(y),\end{gathered} \tag{19}$$

$$Frame(Color(x, c)), \tag{20}$$

$$\begin{aligned}Holds(Color(x, c_1), s) \wedge Holds(Color(x, c_2), s)\\ \supset c_1 = c_2,\end{aligned} \tag{21}$$

$$\begin{gathered}Paint(x_1, c_1) = Paint(x_2, c_2) \supset x_1 = x_2 \wedge c_1 = c_2,\\ Paint(x, c) \neq Move(y, z),\end{gathered} \tag{22}$$

$$\begin{aligned}&Possible(Paint(x, c), s)\\ &\quad \supset Holds(Color(x, c), Result(Paint(x, c), s)),\end{aligned} \tag{23}$$

$$Possible(Paint(x, c), s). \tag{24}$$

The extended theory satisfies the conditions of the main theorem, if the additional axioms are treated in the following way. Formulas (19) and (22) are included in the set of inessential axioms. The universal closure of (20) is appended to F as another conjunctive term. The axiom (21) is replaced by adding

$$\forall x c_1 c_2 [\sigma(Color(x, c_1)) \wedge \sigma(Color(x, c_2)) \supset c_1 = c_2]$$

to $C_0(\sigma)$. Instead of including (23), we disjunctively append

$$\exists xc(a = Paint(x, c) \wedge f = Color(x, c))$$

[6]We argue in [Gelfond *et al.*, 1991] that these ideas can be conveniently expressed in the language of the situation calculus, so that its attractive syntax should not be necessarily tied to the primitive ontology of action accepted in this paper.

to $R_0(a,f)$. Finally, instead of including (24), we disjunctively append

$$\exists xc(a = Paint(x,c))$$

to $P_0(a,\sigma)$.

The characterization of *Noninertial* given by the main theorem will allow us to prove, for instance, that moving blocks does not change their colors, and painting blocks does not change their locations.

4 PROOF OF THE MAIN THEOREM

The theorem follows from three lemmas, given in Section 4.3–4.5. Sections 4.1 and 4.2 contain some preliminary results.

4.1 A LEMMA ABOUT CIRCUMSCRIPTION

We will need the following general property of circumscription $\mathrm{CIRC}[A(P,Z);P;Z]$, which generalizes Theorem 6.4 from [Genesereth and Nilsson, 1987].

Lemma 1. *Let E be a predicate expression without parameters, containing neither P nor Z. If the sentences*

$$A(P,Z) \supset \exists z A(E,z) \tag{25}$$

and

$$A(P,Z) \supset E \leq P \tag{26}$$

are universally valid, then so is the sentence

$$\mathrm{CIRC}[A(P,Z);P;Z] \equiv A(P,Z) \wedge P = E. \tag{27}$$

Proof. To prove (27) left to right, assume $\mathrm{CIRC}[A(P,Z);P;Z]$, that is,

$$A(P,Z) \wedge \neg\exists pz[A(p,z) \wedge p < P].$$

From the first conjunctive term we conclude, using (26), that $E \leq P$, and, using (25), that $\exists z A(E,z)$. From the second conjunctive term,

$$\neg\exists z[A(E,z) \wedge E < P].$$

The last two formulas imply $\neg(E < P)$. In combination with $E \leq P$, this gives $P = E$. Right to left: assume

$$A(P,Z) \wedge P = E. \tag{28}$$

Since (26) is universally valid, so is

$$\forall pz[A(p,z) \supset E \leq p] \tag{29}$$

(constants in a universally valid formula can be replaced by universally quantified variables). Assume $A(p,z) \wedge p < P$. Then, by (29), $E \leq p < P$, contrary to the second term of (28). This contradiction proves the second term of the circumscription in (27).

By essentially the same argument we can prove the relativized form of Lemma 1: For any set Γ of sentences containing neither P nor Z, if Γ implies (25) and (26) (that is, if these sentences are true in every model of Γ), then Γ implies (27).

4.2 THE FUNCTION ρ

In the remaining part of Section 4, T is a theory satisfying Conditions A and B. By A we denote the conjunction of all axioms of T, and by A_0 the conjunction of the inessential axioms, so that

$$A \equiv A_0 \wedge F \wedge S \wedge U \wedge C \wedge R \wedge P \wedge I.$$

Define the function ρ by:

$$\rho[a,\sigma] = \lambda f.[\sigma(f) \vee R_0(a,f)] \\ \wedge \forall f_1[R_0(a,f_1) \supset Compatible(f,f_1)].$$

Intuitively, this is the counterpart of the function *Result* which works on states instead of situations.

The following two lemmas summarize the properties of ρ used in the proof of the theorem.

Lemma 2. *A_0 and F imply*

$$IsState(\sigma) \supset IsState(\rho[a,\sigma]), \tag{30}$$

$$C_0(\sigma) \supset C_0(\rho[a,\sigma]), \tag{31}$$

and

$$R_0(a,f) \supset \rho[a,\sigma](f). \tag{32}$$

Proof. The fact that A_0 and F imply (30) follows from Condition A. To prove (31), assume A_0 and $C_0(\sigma)$. Then, according to part 2 of Condition B,

$$\forall f_1 f_2[\sigma(f_1) \wedge \sigma(f_2) \supset Compatible(f_1,f_2)]. \tag{33}$$

We need to prove $C_0(\rho[a,\sigma])$, that is,

$$\forall f_1 f_2[\rho[a,\sigma](f_1) \wedge \rho[a,\sigma](f_2) \supset Compatible(f_1,f_2)].$$

Assume $\rho[a,\sigma](f_1)$ and $\rho[a,\sigma](f_2)$. Then, by the definition of ρ,

$$\sigma(f_1) \vee R_0(a,f_1), \tag{34}$$

$$R_0(a,f_2) \supset Compatible(f_1,f_2), \tag{35}$$

$$\sigma(f_2) \vee R_0(a,f_2), \tag{36}$$

$$R_0(a,f_1) \supset Compatible(f_2,f_1). \tag{37}$$

Our goal is to derive $Compatible(f_1,f_2)$. If $\sigma(f_1)$ and $\sigma(f_2)$, then this conclusion follows by (33). If not, then, by (34) and (36), $R_0(a,f_1)$ or $R_0(a,f_2)$. In the first case, use (37) and part 1 of Condition B.¡ In the second case, use (35). Formula (31) is proved. To prove (32), assume $R_0(a,f)$. According to the definition of ρ, we need only to check that $R_0(a,f_1) \supset Compatible(f,f_1)$; this follows from part 3 of Condition B.

Lemma 3. *A_0, F and C imply*

$$IsState(\rho[a,State[s]]) \wedge C_0(\rho[a,State[s]]). \tag{38}$$

Proof: (38) follows from (30), (31) and C.

4.3 THE EFFECT OF CIRCUMSCRIBING *Frame*

Lemma 4. *The circumscription*

$$\mathrm{CIRC}[A; \mathit{Frame}; \mathit{Inconsistent}, \mathit{Noninertial}] \tag{39}$$

is equivalent to the conjunction of A and (14).

Proof. By Proposition 2 from [Lifschitz, 1985],

$$\begin{aligned}
&\mathrm{CIRC}[A; \mathit{Frame}; \mathit{Inconsistent}, \mathit{Noninertial}]\\
&\equiv A \wedge \mathrm{CIRC}[(\exists \mathit{Inconsistent}, \mathit{Noninertial}.A);\\
&\qquad\qquad \mathit{Frame}]\\
&\equiv A \wedge \mathrm{CIRC}[A_0 \wedge F \wedge (\exists \mathit{Inconsistent}.S)\\
&\qquad \wedge U \wedge C \wedge R \wedge P \wedge (\exists \mathit{Noninertial}.I); \mathit{Frame}]\\
&\equiv A \wedge \mathrm{CIRC}[A_0 \wedge F \wedge U \wedge C \wedge R \wedge P; \mathit{Frame}].
\end{aligned}$$

The formulas A_0, R and P do not contain *Frame*. Both occurrences of *Frame* in U (as parts of $\mathit{IsState}[\sigma_1]$ and $\mathit{IsState}[\sigma_2]$) are negative. According to part 2 of Condition B, C is equivalent, in the presence of A_0, to a formula not containing *Frame*. Consequently, the conjunction $A_0 \wedge U \wedge C \wedge R \wedge P$ is equivalent to a formula negative relative to *Frame*. By the lemma from Section 4 of [Lifschitz, 1987], it follows that

$$\begin{aligned}
&\mathrm{CIRC}[A_0 \wedge F \wedge U \wedge C \wedge R \wedge P; \mathit{Frame}]\\
&\equiv \mathrm{CIRC}[F; \mathit{Frame}] \wedge A_0 \wedge U \wedge R \wedge C \wedge P.
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathrm{CIRC}[A; \mathit{Frame}; \mathit{Inconsistent}, \mathit{Noninertial}]\\
&\equiv A \wedge \mathrm{CIRC}[F; \mathit{Frame}].
\end{aligned}$$

4.4 THE EFFECT OF CIRCUMSCRIBING *Inconsistent*

Lemma 5. *The circumscription*

$$\begin{aligned}
&\mathrm{CIRC}[A; \mathit{Inconsistent}; \mathit{Holds}, \mathit{Result},\\
&\qquad\qquad \mathit{Possible}, \mathit{Noninertial}]
\end{aligned} \tag{40}$$

is equivalent to the conjunction of A and (16).

Proof. Notice first that (40) is equivalent to

$$\begin{aligned}
&A \wedge \mathrm{CIRC}[\exists \mathit{Possible}, \mathit{Noninertial}.A;\\
&\qquad\qquad \mathit{Inconsistent}; \mathit{Holds}, \mathit{Result}]
\end{aligned}$$

(apply Proposition 2 from [Lifschitz, 1985] to both circumscriptions). The formula

$$\exists \mathit{Possible}, \mathit{Noninertial}.A$$

can be simplified as follows:

$$\begin{aligned}
&\exists \mathit{Possible}, \mathit{Noninertial}.A\\
&\equiv A_0 \wedge F \wedge S \wedge U \wedge C \wedge (\exists \mathit{Possible}.R \wedge P)\\
&\qquad\qquad \wedge(\exists \mathit{Noninertial}.I)\\
&\equiv A_0 \wedge F \wedge S \wedge U \wedge C \wedge (\exists \mathit{Possible}.R \wedge P)\\
&\equiv A_0 \wedge F \wedge S \wedge U \wedge C \wedge R^*,
\end{aligned}$$

where R^* stands for

$$P_0(a, \mathit{State}[s]) \wedge R_0(a, f) \supset \mathit{Holds}(f, \mathit{Result}(a, s)).$$

It follows that (40) is equivalent to

$$\begin{aligned}
&A \wedge \mathrm{CIRC}[A_0 \wedge F \wedge S \wedge U \wedge C \wedge R^*;\\
&\qquad\qquad \mathit{Inconsistent}; \mathit{Holds}, \mathit{Result}].
\end{aligned}$$

We will compute this circumscription using Lemma 1, with the expression

$$\lambda\sigma[\mathit{IsState}(\sigma) \wedge \neg C_0(\sigma)]$$

as E. We need to prove

$$\begin{aligned}
&A_0 \wedge F \wedge S(\mathit{Inconsistent}, \mathit{Holds}) \wedge U\\
&\quad \wedge C(\mathit{Holds}) \wedge R^*(\mathit{Holds}, \mathit{Result})\\
&\quad\quad \supset \exists \mathit{holds}, \mathit{result}[A_0 \wedge F \wedge S(E, \mathit{holds})\\
&\qquad\qquad \wedge U \wedge C(\mathit{holds}) \wedge R^*(\mathit{holds}, \mathit{result})]
\end{aligned}$$

and

$$\begin{aligned}
&A_0 \wedge F \wedge S(\mathit{Inconsistent}, \mathit{Holds}) \wedge U\\
&\quad \wedge C(\mathit{Holds}) \wedge R^*(\mathit{Holds}, \mathit{Result})\\
&\qquad \supset E \le \mathit{Inconsistent}.
\end{aligned}$$

In other words, assuming

$$\begin{aligned}
&A_0 \wedge F \wedge S(\mathit{Inconsistent}, \mathit{Holds}) \wedge U\\
&\quad \wedge C(\mathit{Holds}) \wedge R^*(\mathit{Holds}, \mathit{Result}),
\end{aligned}$$

we need to select *holds* and *result*, and prove:

$$\begin{gathered}
S(E, \mathit{holds}),\\
C(\mathit{holds}),\\
R^*(\mathit{holds}, \mathit{result}),\\
E \le \mathit{Inconsistent}.
\end{gathered}$$

The predicate *holds* is defined by cases, depending on whether or not there is a σ satisfying the condition

$$\mathit{IsState}(\sigma) \wedge C_0(\sigma) \wedge s = \mathit{Sit}(\sigma) \tag{41}$$

$$\mathit{holds}(f, s) \equiv \begin{cases} \sigma(f), & \text{if } \sigma \text{ satisfies (41)},\\ \mathit{false}, & \text{if there is no such } \sigma. \end{cases}$$

The correctness of this definition follows from the fact that, by U, there can be at most one σ satisfying (41). The function *result* is defined by

$$\mathit{result}(a, s) = \mathit{Sit}(\rho[a, \mathit{State}[s]]).$$

The formula $S(E, \mathit{holds})$, that is,

$$\begin{aligned}
&\mathit{IsState}(\sigma) \wedge \neg[\mathit{IsState}[\sigma] \wedge \neg C_0(\sigma)]\\
&\supset \sigma = \lambda f[\mathit{Frame}(f) \wedge \mathit{holds}(f, \mathit{Sit}(\sigma))],
\end{aligned}$$

is propositionally equivalent to

$$\begin{aligned}
&\mathit{IsState}(\sigma) \wedge C_0(\sigma)\\
&\supset \sigma = \lambda f[\mathit{Frame}(f) \wedge \mathit{holds}(f, \mathit{Sit}(\sigma))].
\end{aligned}$$

Assume $\mathit{IsState}(\sigma)$ and $C_0(\sigma)$. Then, using the definition of *holds*, we compute:

$$\begin{aligned}
&\lambda f[\mathit{Frame}(f) \wedge \mathit{holds}(f, \mathit{Sit}(\sigma))]\\
&= \lambda f[\mathit{Frame}(f) \wedge \sigma(f)]\\
&= \lambda f.\sigma(f) = \sigma.
\end{aligned}$$

The formula $C(\mathit{holds})$ can be written as

$$C_0(\lambda f[\mathit{Frame}(f) \wedge \mathit{holds}(f, s)]).$$

If σ is such that $IsState(\sigma) \wedge C_0(\sigma) \wedge s = Sit(\sigma)$, then, according to the definition of *holds*, $holds(f, s) \equiv \sigma(f)$, so that $C(holds)$ is equivalent to

$$C_0(\lambda f[Frame(f) \wedge \sigma(f)]).$$

Since $IsState(\sigma)$, this is the same as $C_0(\sigma)$. On the other hand, if there is no σ satisfying the condition $IsState(\sigma) \wedge C_0(\sigma) \wedge s = Sit(\sigma)$, then $C(holds)$ becomes $C_0(\lambda f.false)$, which follows from part 2 of Condition B.

The formula $R^*(holds, result)$ can be written as

$$\begin{aligned} &P_0(a, \lambda f(Frame(f) \wedge holds(f, s))) \\ &\quad \wedge R_0(a, f) \supset holds(f, result(a, s)). \end{aligned}$$

We can actually prove the stronger formula

$$R_0(a, f) \supset holds(f, result(a, s)),$$

that is,

$$R_0(a, f) \supset holds(f, Sit(\rho[a, State[s]])).$$

Using (38) and the definition of *holds*, this can be further rewritten as

$$R_0(a, f) \supset \rho[a, State[s]](f),$$

which follows from (32).

Finally, $E \leq Inconsistent$ stands for

$$IsState(\sigma) \wedge \neg C_0(\sigma) \supset Inconsistent(\sigma).$$

Assume $IsState(\sigma)$ and $\neg Inconsistent(\sigma)$. It follows then from S that $\sigma = State[s]$ for some s. Then $C_0(State[s])$ follows from C.

4.5 THE EFFECT OF CIRCUMSCRIBING *Noninertial*

Lemma 6. *Assuming (15), the circumscription*

$$\text{CIRC}[A; Noninertial; Result] \tag{42}$$

is equivalent to the conjunction of A and (16).

Proof. We will use the relativized form of Lemma 1, with (15) as Γ and the expression

$$\begin{aligned} &\lambda fas[Frame(f) \wedge Possible(a, s) \\ &\quad \wedge Affected(f, a, State[s])] \end{aligned}$$

as E. We need to prove that (15) implies

$$A(Noninertial, Result) \supset \exists result\, A(E, result)$$

and

$$A(Noninertial, Result) \supset E \leq Noninertial.$$

Assume $A(Noninertial, Result)$, and define *result* as in the proof of Lemma 5. Since the only parts of A that contain *Noninertial* or *Result* are R and I, we need only to prove:

$$\begin{gathered} R(result), \\ I(E, Result), \\ E \leq Noninertial. \end{gathered}$$

The following relationship between *result* and ρ will be used in the proof:

$$State[result(a, s)] = \rho[a, State[s]]. \tag{43}$$

To prove this, notice that (38) and (15) imply $\neg Inconsistent(\rho[a, State[s]])$. Using (38) and S, we conclude that the right-hand side of (43) equals $State[Sit(\rho[a, State[s]])]$. By the definition of *result*, this is the same as the right-hand side of (43).

The formula $R(result)$ is

$$Possible(a, s) \wedge R_0(a, f) \supset Holds(f, result(a, s)).$$

We will prove a stronger formula:

$$R_0(a, f) \supset Holds(f, result(a, s)).$$

Assume $R_0(a, f)$. Then, by (32), $\rho[a, State[s]](f)$. It follows by (43) that $State[result(a, s)](f)$, and consequently $Holds(f, result(a, s))$.

The formula $I(E, holds)$ is

$$\begin{aligned} &Frame(f) \wedge Possible(a, s) \wedge \neg E(f, a, s) \\ &\supset [Holds(f, result(a, s) \equiv Holds(f, s)], \end{aligned}$$

which is propositionally equivalent to

$$\begin{aligned} &Frame(f) \\ &\quad \wedge Possible(a, s) \\ &\qquad \wedge \neg Affected(f, a, State[s]) \\ &\qquad\qquad \supset [Holds(f, result(a, s)) \equiv Holds(f, s)]. \end{aligned}$$

Assume that $Frame(f)$ and $Possible(a, s)$, but

$$\neg[Holds(f, result(a, s)) \equiv Holds(f, s)]. \tag{44}$$

Using (43), we can rewrite this in the form

$$\neg\{\rho[a, State[s]](f) \equiv State[s](f)\},$$

which, by the definition of ρ, is the same as

$$\begin{aligned} &\neg\{\{[(State[s](f) \vee R_0(a, f)] \\ &\quad \wedge \forall f_1[R_0(a, f_1) \supset Compatible(f, f_1)]\} \\ &\quad \equiv State[s](f)\}. \end{aligned} \tag{45}$$

Our goal is to prove

$$Affected(f, a, State[s]),$$

that is,

$$\begin{aligned} &IsState(\sigma_1) \wedge C_0(\sigma_1) \wedge [\sigma_1(f) \equiv State[s](f)] \\ &\qquad \supset \exists f_1[R_0(a, f_1) \wedge \neg\sigma_1(f_1)]. \end{aligned} \tag{46}$$

Assume $IsState(\sigma_1)$, $C_0(\sigma_1)$ and

$$\sigma_1(f) \equiv State[s](f).$$

Using the last equivalence $\sigma_1(f) \equiv State[s](f)$, we can rewrite (45) in the form

$$\begin{aligned} &\neg\{\{[(\sigma_1(f) \vee R_0(a, f)] \\ &\wedge \forall f_1[R_0(a, f_1) \supset Compatible(f, f_1)]\} \equiv \sigma_1(f)\}. \end{aligned} \tag{47}$$

Case 1: $R_0(a, f)$. Then, by part 3 of Condition B,

$$\forall f_1[R_0(a, f_1) \supset Compatible(f, f_1)],$$

so that the left-hand side of the equivalence in (47) is true, and (47) implies $\neg\sigma_1(f)$. We see that the consequent of (46) is true for $f_1 = f$. Case 2: $\neg R_0(a, f)$. Then (47) can be rewritten as

$$\neg\{\{(\sigma_1(f) \wedge \forall f_1[R_0(a, f_1) \supset Compatible(f, f_1)]\} \equiv \sigma_1(f)\}.$$

This is only possible when $\sigma_1(f)$ and

$$\neg\forall f_1[R_0(a, f_1) \supset Compatible(f, f_1)].$$

Take f_1 such that $R_0(a, f_1)$ and $\neg Compatible(f, f_1)$. By part 2 of Condition B,

$$\sigma_1(f) \wedge \sigma_1(f_1) \supset Compatible(f, f_1).$$

Consequently, $\neg\sigma_1(f_1)$, which again proves the consequent of (46).

It remains to show that $E \leq Noninertial$, that is,

$$Frame(f) \wedge Possible(a, s) \wedge Affected(f, a, State[s]) \supset Noninertial(f, a, s).$$

We will assume $Frame(f)$, $Possible(a, s)$ and

$$\neg Noninertial(f, a, s),$$

and prove

$$\neg Affected(f, a, State[s]),$$

that is,

$$\exists\sigma_1[IsState(\sigma_1) \wedge C_0(\sigma_1) \wedge (\sigma_1(f) \equiv State[s](f) \wedge\neg\exists f_1[R_0(a, f_1) \wedge \neg\sigma_1(f_1)]].$$

By I,

$$Holds(f, Result(a, s)) \equiv Holds(f, s). \quad (48)$$

Let us check that $\sigma_1 = State[Result(a, s)]$ has all required properties. Clearly, $IsState(\sigma_1)$; C implies $C_0(\sigma_1)$; (48) implies $\sigma_1(f) \equiv State[s](f)$. Let f_1 be such that $R_0(a, f_1)$. Then, by R, $Holds(f_1, Result(a, s))$, that is, $\sigma_1(f_1)$.

Acknowledgments

The author would like to thank Michael Gelfond, Fangzhen Lin, John McCarthy and Raymond Reiter for useful discussions on the subject of this paper and for comments on an earlier draft. He is also grateful to the anonymous referees for their comments on the abstract of this paper. The work on the paper began when the author worked at Stanford University and was supported in part by NSF grant IRI-89-04611 and by DARPA under Contract N00039-84-C-0211.

References

[Baker and Ginsberg, 1989] Andrew Baker and Matthew Ginsberg. Temporal projection and explanation. In *Proc. of IJCAI-89*, pages 906–911, 1989.

[Baker, 1989] Andrew Baker. A simple solution to the Yale Shooting Problem. In Ronald Brachman, Hector Levesque, and Raymond Reiter, editors, *Proc. of the First Int'l Conf. on Principles of Knowledge Representation and Reasoning*, pages 11–20, 1989.

[Gelfond *et al.*, 1991] Michael Gelfond, Vladimir Lifschitz, and Arkady Rabinov. What are the limitations of the situation calculus? In Robert Boyer, editor, *Automated Reasoning: Essays in Honor of Woody Bledsoe*, pages 167–179. Kluwer Academic, Dordrecht, 1991.

[Genesereth and Nilsson, 1987] Michael R. Genesereth and Nils J. Nilsson. *Logical Foundations of Artificial Intelligence.* Morgan Kaufmann, San Mateo, CA, 1987.

[Lifschitz, 1985] Vladimir Lifschitz. Computing circumscription. In *Proc. of IJCAI-85*, pages 121–127, 1985.

[Lifschitz, 1987] Vladimir Lifschitz. Pointwise circumscription. In Matthew Ginsberg, editor, *Readings in nonmonotonic reasoning*, pages 179–193. Morgan Kaufmann, San Mateo, CA, 1987.

[Lifschitz, 1990] Vladimir Lifschitz. Frames in the space of situations. *Artificial Intelligence*, 46:365–376, 1990.

[McCarthy and Hayes, 1969] John McCarthy and Patrick Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence*, volume 4, pages 463–502. Edinburgh University Press, Edinburgh, 1969. Reproduced in [McCarthy, 1990].

[McCarthy, 1986] John McCarthy. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence*, 26(3):89–116, 1986. Reproduced in [McCarthy, 1990].

[McCarthy, 1990] John McCarthy. *Formalizing common sense: papers by John McCarthy.* Ablex, Norwood, NJ, 1990.