Situation Calculus and Causal Logic

Vladimir Lifschitz
Department of Computer Sciences
University of Texas at Austin
Austin, TX 78712, USA
vl@cs.utexas.edu

Abstract

In the AAAI-97 paper by McCain and Turner, the frame problem is solved using a nonmonotonic causal logic. In this note we show how their method can be adapted to the language of the situation calculus. The "causal situation calculus" is applied to action domains involving ramifications and nondeterminism, and related to the causal formalization of actions proposed by Lin.

1 Introduction

In this note, a nonmonotonic causal logic is used for describing actions in the situation calculus. This logic was defined for the propositional case in [McCain and Turner, 1997] on the basis of earlier work by several authors, including [Geffner, 1990], [Lin, 1995] and [McCain and Turner, 1995]. In [Lifschitz, 1997], it is extended to languages with variables. It is closely related to special cases of "general purpose" nonmonotonic formalisms such as logic programming, default logic and circumscription [Turner, 1998], and appears to be particularly well suited for the task of describing the effects of actions.

In [McCain and Turner, 1997], values of fluents are represented by propositional atoms indexed by time. For instance, the atom Closed_5 would express that the fluent *Closed* holds at time 5. Actions are treated in a similar way: the atom Close_5 would express that the action *Close* was performed between time instants 5 and 6. Action descriptions in [Lifschitz, 1997] follow the same pattern, except that when variables are available in the language, a fluent can be represented by a predicate symbol that takes time as the argument, or as one of several arguments. We can express, for instance, that the action of moving an object x to a location l was performed between times t and $t+1$ by the atomic formula $\mathit{Move}(x, l, t)$.

This "time-based" representation is different from the situation calculus approach [McCarthy and Hayes, 1969], which represents actions by terms and would use an expression like $\mathit{Result}(\mathit{Move}(x, l), s)$ to denote the situation that results from moving x to location l in situation s. Advantages and disadvantages of the ontology and syntax of the situation calculus have been the subject of many debates in the knowledge representation community. It is clear that at least in simple cases the language of the situation calculus is very attractive.

The purpose of this note is to show that the version of the nonmonotonic causal logic presented in [Lifschitz, 1997] is applicable in the framework of the situation calculus as well, that the elegant solution to the frame problem proposed in [McCain and Turner, 1997] can be easily expressed in the language of the situation calculus, and that the process of "literal completion" described in these papers may allow us to translate this solution into first-order logic.

To facilitate the comparison of the causal version of the situation calculus with earlier work, we apply it here to a formalization of the blocks world which is similar to the one presented in [Lifschitz, 1991]. This formalization involves actions with indirect effects. Then we discuss an enhancement of the basic example in which the effects of actions can be nondeterministic.

The description of the blocks world used in [Lifschitz, 1991] is outlined in Section 2. In Sections 3 and 4 we turn this description into a causal theory and show how this theory can be converted into a set of first-order axioms. The relationship between our method and Lin's is discussed in Section 5. The enhancement in which nondeterministic actions are included is presented in Section 6.

To make this paper self-contained, we reproduce the definition of the semantics of causal theories from [Lifschitz, 1997] in Appendix A. Appendix B contains proofs of theorems.

2 Indirect Effects of Moving a Block

We will describe the locations of blocks using the fluent expressions

$$On(x, y)$$

(block x is on block y) and

$$OnT(x)$$

(block x is on the table), which will be declared "inertial," that is, subject to the commonsense law of inertia. Fluent expressions of the form

$$Clear(x)$$

will be characterized by relating them to On, and they will not be inertial.

The effect of placing a block x on top of block y is described in [Lifschitz, 1991] by the axiom[1]

$$\begin{array}{l} Poss(Move(x, y), s) \\ \quad \supset Holds(On(x, y), Result(Move(x, y), s)). \end{array} \tag{1}$$

On the other hand, a block cannot be in two places at once:

$$\begin{array}{l} Holds(On(x, y), s) \wedge Holds(On(x, z), s) \supset y = z, \\ \neg(Holds(On(x, y), s) \wedge Holds(OnT(x), s)). \end{array} \tag{2}$$

In view of this fact, moving a block may have an "indirect" effect on an inertial fluent: performing the action $Move(x, y)$ will make $On(x, z)$ false if z is the current location of x and $y \neq z$, and it will make $OnT(x)$ false if x is currently on the table.

The presence of indirect effects makes the frame problem more difficult. (This aspect of the frame problem is known as the "ramification problem.") In [Lifschitz, 1991], this difficulty is addressed using circumscription in the style of [Baker, 1991]. The logical machinery that is employed there is rather complicated. It includes a second-order formulation of the existence of situation axioms and conjoining three circumscriptions with carefully selected lists of varied constants. In this note we show that the same goal can be achieved more easily if the effects of moving a block are described by causal rules.

3 A Causal Theory of the Blocks World

In causal logic, we distinguish between the claim that a proposition is true and the stronger claim that there is a *cause* for it to be true.

Consider a language of classical logic, possibly many-sorted or higher-order. A *causal rule* is an expression of the form

$$F \leftarrow G \tag{3}$$

where F and G are formulas of this language, called the *head* and the *body* of the rule. Intuitively, (3) expresses that F has a cause if G is true, or that G provides a "causal explanation" for F. If the body G of (3) is the logical constant *True*, we drop it and write the rule as $F \leftarrow$. A formula F is identified with the rule

$$False \leftarrow \neg F. \tag{4}$$

A *causal theory* is defined by

- a finite set of nonlogical constants[2] of the underlying language, called the *explainable symbols* of the theory, and
- a finite set of causal rules.

The semantics of causal theories is defined in [Lifschitz, 1997] by a translation into classical logic; a causal theory is viewed as shorthand for the sentence obtained as the result of this translation. In particular, a *model* of a causal theory T is a model of its translation, in the sense of classical logic; a *theorem* of T is a sentence that is logically entailed by the translation of T. The definition of the translation is reproduced in Appendix A. Its main feature is that it can be written as the conjunction of the universal closures of the implications

$$G \supset F \tag{5}$$

for all rules (3) of the theory plus an additional conjunctive term. (This term makes the translation nonmonotonic, just like the minimality condition in the definition of circumscription.) In this sense, a causal rule (3) is stronger than the corresponding implication (5).

We will define now a causal theory of the blocks world, CBW. The explainable symbols of the theory are the predicate constants *Holds* and *Inertial*.

Formula (1) is replaced in CBW by the rule

$$\begin{array}{r} Holds(On(x, y), Result(Move(x, y), s)) \\ \leftarrow Poss(Move(x, y), s). \end{array} \tag{6}$$

Thus instead of postulating merely that the fluent $On(x, y)$ holds after performing $Move(x, y)$, we say that there is a *cause* for this fluent to hold.

[1] The action of putting a block on the table is not part of this formalization, but it can be easily added.

[2] A *nonlogical constant* is a function constant or a predicate constant (other than equality).

Formula (2) is replaced by the rules saying that, wherever a block is located, there is a cause for it not to be anywhere else:

$$\begin{aligned}\neg Holds(On(x,z),s) &\leftarrow Holds(On(x,y),s) \wedge y \neq z,\\ \neg Holds(OnT(x),s) &\leftarrow Holds(On(x,y),s),\\ \neg Holds(On(x,y),s) &\leftarrow Holds(OnT(x),s).\end{aligned} \tag{7}$$

The predicate constant *Inertial* is characterized by the rules

$$\begin{aligned}Inertial(On(x,y)) &\leftarrow ,\\ Inertial(OnT(x)) &\leftarrow ,\\ \neg Inertial(f) &\leftarrow \neg Inertial(f).\end{aligned} \tag{8}$$

The last rule says that if a fluent is not inertial then there is a cause for this. Proposition 1 from [Lifschitz, 1997] shows that including this rule has the same effect as circumscribing *Inertial*.

In accordance with a proposal from [McCain and Turner, 1997], the commonsense law of inertia is understood as the assertion that if an inertial fluent does not change its value after an action is executed then there is a cause for this:

$$\begin{aligned}Holds(f,Result(a,s)) \leftarrow\ & Inertial(f) \wedge Poss(a,s)\\ & \wedge Holds(f,s)\\ & \wedge Holds(f,Result(a,s)),\\ \neg Holds(f,Result(a,s)) \leftarrow\ & Inertial(f) \wedge Poss(a,s)\\ & \wedge \neg Holds(f,s)\\ & \wedge \neg Holds(f,Result(a,s)).\end{aligned} \tag{9}$$

Rules (6), (7) and (9) provide causal explanations for the truth values of $Holds(f,s)$ only when f is inertial and s is the result of performing an action. In all other cases, we postulate that, whatever the truth value of $Holds(f,s)$ is, there is a cause for this:

$$\begin{aligned}Holds(f,s) &\leftarrow Holds(f,s) \wedge \neg Inertial(f),\\ \neg Holds(f,s) &\leftarrow \neg Holds(f,s) \wedge \neg Inertial(f),\\ Holds(f,s) &\leftarrow Holds(f,s) \wedge \neg\exists a s_1 (Poss(a,s_1)\\ &\qquad \wedge s = Result(a,s_1)),\\ \neg Holds(f,s) &\leftarrow \neg Holds(f,s) \wedge \neg\exists a s_1 (Poss(a,s_1)\\ &\qquad \wedge s = Result(a,s_1)).\end{aligned} \tag{10}$$

In addition to these causal rules, we need a few postulates that are expressed by formulas. (Recall that a formula can be viewed as a causal rule (4) whose head is the logical constant *False*.) One of these formulas tells us when the action $Move(x,y)$ is executable:

$$\begin{aligned}Poss(Move(x,y),s) \equiv\ & Holds(Clear(x),s)\\ & \wedge Holds(Clear(y),s)\\ & \wedge x \neq y.\end{aligned} \tag{11}$$

We include also the definition of *Clear* in terms of *On*:

$$Holds(Clear(x),s) \equiv \neg\exists y Holds(On(y,x),s). \tag{12}$$

The remaining postulates express the unique name assumption. The terms $On(x,y)$, $OnT(x)$ and $Clear(x)$ represent distinct fluents:

$$\begin{gathered}On(x_1,y_1) = On(x_2,y_2) \supset x_1 = x_2 \wedge y_1 = y_2,\\ OnT(x) = OnT(y) \supset x = y,\\ Clear(x) = Clear(y) \supset x = y,\\ On(x,y) \neq OnT(z),\\ On(x,y) \neq Clear(z),\\ OnT(x) \neq Clear(y).\end{gathered} \tag{13}$$

The terms $Move(x,y)$ represent distinct actions:

$$Move(x_1,y_1) = Move(x_2,y_2) \supset x_1 = x_2 \wedge y_1 = y_2. \tag{14}$$

A situation contains complete information about its past:

$$Result(a_1,s_1) = Result(a_2,s_2) \supset a_1 = a_2 \wedge s_1 = s_2. \tag{15}$$

This is the end of the list of rules of *CBW*.

4 Simplifying *CBW*

According to the semantics of causal theories, *CBW* is shorthand for a sentence of classical logic that includes second-order quantifiers. The process of literal completion allows us to rewrite this sentence in first-order logic:

Proposition 1. *CBW is equivalent to the conjunction of the universal closures of formulas (1), (2), (11)–(15),*

$$\begin{aligned}& Inertial(f)\\ &\quad \equiv \exists xy(f = On(x,y)) \vee \exists x(f = OnT(x)),\end{aligned} \tag{16}$$

$$\begin{aligned}& Inertial(f) \wedge Poss(a,s)\\ &\quad \wedge \neg Holds(f,s) \wedge Holds(f,Result(a,s))\\ &\qquad \supset \exists xy[f = On(x,y) \wedge a = Move(x,y)]\end{aligned} \tag{17}$$

and

$$\begin{aligned}& Inertial(f) \wedge Poss(a,s)\\ & \wedge Holds(f,s) \wedge \neg Holds(f,Result(a,s))\\ & \supset \exists xyz[f = On(x,z) \wedge Holds(On(x,y),Result(a,s))\\ &\qquad \wedge y \neq z]\\ & \vee \exists xy[f = OnT(x) \wedge Holds(On(x,y),Result(a,s))]\\ & \vee \exists xy[f = On(x,y) \wedge Holds(OnT(x),Result(a,s))].\end{aligned} \tag{18}$$

Formula (16) is an explicit definition of *Inertial*. Formula (17) is an explanation closure axiom in the sense of [Schubert, 1990] and [Reiter, 1991]: it tells us under what condition a fluent f can change its value from *false* to *true* after executing an action. It says that, for an inertial fluent, this can only happen if the fluent has the form $On(x, y)$, and if the action that has been executed is $Move(x, y)$.

Formula (18) can be viewed as a second explanation closure axiom—it tells us under what condition f can change its value from *true* to *false*. But it does this in a somewhat indirect way, without actually saying which action has been executed. We can be more explicit: an inertial fluent can change its value from *true* to *false* after executing an action only if

- the fluent has the form $On(x, y)$, and the action consisted in placing x on top of a block other than y, or
- the fluent has the form $OnT(x)$, and the action consisted in placing x on top of a block.

The following corollary to Proposition 1 shows that this is indeed an equivalent reformulation of (18):

Corollary. *CBW is equivalent to the conjunction of the universal closures of formulas (1), (2), (11)–(17) and*

$$\begin{aligned}&Inertial(f) \wedge Poss(a, s)\\ &\wedge Holds(f, s) \wedge \neg Holds(f, Result(a, s))\\ &\quad \supset \exists xyz[f = On(x, y) \wedge a = Move(x, z) \wedge y \neq z]\\ &\qquad \vee \exists xy[f = OnT(x) \wedge a = Move(x, y)].\end{aligned} \tag{19}$$

5 Comparison with Lin's Method

Lin [1995] expresses the basic distinction between being true and having a cause without introducing a specialized nonmonotonic formalism. Instead, he extends the situation calculus language by the new predicate *Caused*, and applies circumscription to minimize this predicate. In this section, we define a Lin-style formalization of the blocks world and relate it to the causal theory *CBW*.

The predicate *Caused* has three arguments: a fluent f, a truth value v, and a situation s in which there is a cause for f to have the value v. It is convenient to replace *Caused* by two binary predicates corresponding to the two possible values of v: $Caused^+(f, s)$ will express that there is a cause for f to be true in the situation s, and $Caused^-(f, s)$ will express that there is a cause for f to be false.

The formulas corresponding to rules (6) and (7) in this formalism are

$$\begin{aligned}&Poss(Move(x, y), s)\\ &\quad \supset Caused^+(On(x, y), Result(Move(x, y), s))\end{aligned} \tag{20}$$

and

$$\begin{aligned}Holds(On(x, y), s) \wedge y \neq z \supset Caused^-(On(x, z), s),\\ Holds(On(x, y), s) \supset Caused^-(OnT(x), s),\\ Holds(OnT(x), s) \supset Caused^-(On(x, y), s).\end{aligned} \tag{21}$$

The predicate *Inertial* is described by the axioms

$$\begin{aligned}Inertial(On(x, y)),\\ Inertial(OnT(x)).\end{aligned} \tag{22}$$

By *LBW* we denote the parallel circumscription of $Caused^+$, $Caused^-$ and *Inertial* in the conjunction of the universal closures of (20)–(22).

To make the result of this circumscription useful for determining the values of fluents, we need to relate the new predicates $Caused^+$ and $Caused^-$ to *Holds* by the axioms

$$\begin{aligned}Caused^+(f, s) \supset Holds(f, s),\\ Caused^-(f, s) \supset \neg Holds(f, s),\end{aligned} \tag{23}$$

and to add an axiom expressing the commonsense law of inertia:

$$\begin{aligned}&Inertial(f) \wedge Poss(a, s)\\ &\quad \wedge \neg Caused^+(f, Result(a, s))\\ &\quad \wedge \neg Caused^-(f, Result(a, s))\\ &\qquad \supset [Holds(f, Result(a, s)) \equiv Holds(f, s)].\end{aligned} \tag{24}$$

Alternatively, inertia can be stated in the explanation closure format:

$$\begin{aligned}&Inertial(f) \wedge Poss(a, s)\\ &\quad \wedge \neg Holds(f, s) \wedge Holds(f, Result(a, s))\\ &\qquad \supset Caused^+(f, Result(a, s)),\\ &Inertial(f) \wedge Poss(a, s)\\ &\quad \wedge Holds(f, s) \wedge \neg Holds(f, Result(a, s))\\ &\qquad \supset Caused^-(f, Result(a, s)).\end{aligned} \tag{25}$$

It is easy to see that the equivalence between (24) and the conjunction of the formulas (25) is a propositional consequence of (23).

The relationship between this formalization of the blocks world and the causal theory *CBW* presented in Section 3 is described in the following proposition:

Proposition 2. *The conjunction of LBW with the universal closures of (11)–(15), (23) and (25) is equivalent to the conjunction of CBW with the universal closures of*

$$\begin{aligned}&Caused^+(f, s)\\ &\quad \equiv \exists xys'[Poss(Move(x, y), s') \wedge f = On(x, y)\\ &\qquad \wedge s = Result(Move(x, y), s')]\end{aligned} \tag{26}$$

and

$$
\begin{aligned}
&Caused^-(f,s)\\
&\quad\equiv \exists xyz[f = On(x,z) \wedge Holds(On(x,y),s)\\
&\qquad\qquad\qquad\qquad \wedge y \neq z]\\
&\qquad \vee \exists xy[f = OnT(x) \wedge Holds(On(x,y),s)]\\
&\qquad \vee \exists xy[f = On(x,y) \wedge Holds(OnT(x),s)].
\end{aligned}
\tag{27}
$$

Thus the Lin-style causal formalization of the blocks world is equivalent to the definitional (and consequently conservative) extension of CBW obtained by adding the definitions (26), (27) of $Caused^+$ and $Caused^-$.

The two formalizations are also similar in that transforming each of them into an equivalent first-order theory is accomplished by a completion process.

On the other hand, these formalizations differ from each other in their approach to inertia. Formula (24), similar to the formalization of the commonsense law of inertia in [Lin, 1995], treats it monotonically: this formula is included as an additional constraint on top of a circumscriptive formalization of causality. By dropping (24) from the list of axioms we would make the theory weaker. Rules (9), which express inertia in CBW and are inspired by [McCain and Turner, 1997], operate nonmonotonically. By dropping them from the list of rules we would make the theory stronger—it would become inconsistent.

6 Nondeterministic Actions

Imagine that a block can be damaged when it is moved to a new location. The effect of the action $Move(x,y)$ on the new fluent $Damaged(x)$ is nondeterministic: if this fluent is false in a situation s then in the new situation $Result(Move(x,y),s)$ it may become true, but it may also remain false.

To express that a block is *always* damaged when it is moved to a new location, we would write the rule

$$
\begin{aligned}
&Holds(Damaged(x), Result(Move(x,y),s))\\
&\qquad\qquad\qquad \leftarrow Poss(Move(x,y),s)),
\end{aligned}
$$

similar to rule (6). In accordance with the approach to nondeterministic actions proposed in [McCain and Turner, 1997], to express that a block *may be* damaged after performing this action, we append a copy of the head of this rule to the body as another conjunctive term:

$$
\begin{aligned}
&Holds(Damaged(x), Result(Move(x,y),s))\\
&\quad \leftarrow Poss(Move(x,y),s)\\
&\qquad \wedge Holds(Damaged(x), Result(Move(x,y),s)).
\end{aligned}
\tag{28}
$$

Rule (28) says that if x damaged after performing the action then there is a cause for this. The head of the rule is asserted here to be a causally justified possibility.

The enhancement CBW' of the theory CBW from Section 3 in which we can talk about damaged blocks is obtained by adding rule (28), the rule

$$Inertial(Damaged(x)) \leftarrow \tag{29}$$

and the unique name axioms

$$
\begin{aligned}
Damaged(x) = Damaged(y) &\supset x = y,\\
Damaged(x) &\neq On(y,z),\\
Damaged(x) &\neq OnT(y),\\
Damaged(x) &\neq Clear(y).
\end{aligned}
\tag{30}
$$

Proposition 3. *CBW' is equivalent to the conjunction of the universal closures of formulas (1), (2), (11)–(15), (18), (30),*

$$
\begin{aligned}
&Inertial(f)\\
&\quad \equiv \exists xy(f = On(x,y)) \vee \exists x(f = OnT(x))\\
&\qquad\qquad\qquad\qquad \vee \exists x(f = Damaged(x))
\end{aligned}
\tag{31}
$$

and

$$
\begin{aligned}
&Inertial(f) \wedge Poss(a,s)\\
&\quad \wedge \neg Holds(f,s) \wedge Holds(f, Result(a,s))\\
&\qquad \supset \exists xy[f = On(x,y) \wedge a = Move(x,y)]\\
&\qquad\quad \vee \exists xy[f = Damaged(x) \wedge a = Move(x,y)].
\end{aligned}
\tag{32}
$$

The last formula differs from the explanation closure condition (17) by the presence of the disjunctive term that describes how a block can be damaged. This is the main difference between the characterization of CBW in Proposition 1 and the characterization of CBW' in Proposition 3.

7 Future Work

The examples presented above show how the causal situation calculus handles ramifications and nondeterminism. Otherwise, the expressive possibilities of this formalism remain unexplored. One particularly interesting issue is the treatment of qualification constraints in the sense of [Lin and Reiter, 1994].

Methodologies for representing actions can be compared in a precise and general way by proving embedding theorems, such as the results presented in [Kartha, 1993]. Our plans for future work include relating the causal situation calculus to the action language $\mathcal{C}$ from [Giunchiglia and Lifschitz, 1998].

Acknowledgements

I am grateful to Enrico Giunchiglia, Fangzhen Lin, Norman McCain and Hudson Turner for useful comments. This work was partially supported by National Science Foundation under grant IRI-9306751.

Appendix A: Semantics of Causal Theories

In the definition of the semantics of causal theories below, we use the substitution of variables for the explainable symbols in a formula. In connection with this, it is convenient to denote formulas by expressions like $F(E)$, where E is the list of all explainable symbols. Then, for any tuple e of variables that is similar[3] to E, the result of replacing all occurrences of the constants E in $F(E)$ by the variables e can be denoted by $F(e)$.

Consider a causal theory T with the explainable symbols E and the causal rules

$$F_i(E, x^i) \leftarrow G_i(E, x^i) \qquad (i = 1, \ldots),$$

where x^i is the list of all free variables of the i-th rule. Take a tuple e of new variables similar to E. By $T^*(e)$ we denote the formula

$$\bigwedge_i \forall x^i (G_i(E, x^i) \supset F_i(e, x^i)).$$

Note that the occurrences of explainable symbols in the heads are replaced here by variables, and the occurrences in the bodies are not. According to the semantics of causal theories, T is shorthand for the sentence

$$\forall e (T^*(e) \equiv e = E). \tag{33}$$

(The expression $e = E$ stands for the conjunction of the equalities between the members of e and the corresponding members of E.)

Intuitively, the condition $T^*(e)$ expresses that the possible values e of the explainable symbols E are "explained" by the rules of T. Sentence (33) says that the actual values of these symbols are the only ones that are explained by the rules of T.

Formula (33) can be equivalently written as

$$T^*(E) \wedge \forall e (T^*(e) \supset e = E). \tag{34}$$

In this expression, the term $T^*(E)$ is the conjunction of the universal closures of the implications

$$G_i(E, x^i) \supset F_i(E, x^i)$$

corresponding to the rules of T. The second conjunctive term of (34) is what makes a causal rule stronger than the corresponding implication, and what makes the formalism nonmonotonic.

[3] The similarity condition means that (i) e has the same length as E, (ii) if the k-th member of E is a function constant then the k-th member of e is a function variable of the same type, and (iii) if the k-th member of E is a predicate constant then the k-th member of e is a predicate variable of the same type.

Appendix B: Proofs of Theorems

B.1 Literal Completion

As defined in [Lifschitz, 1997], a causal theory T is *definite* if every explainable symbol of T is a predicate constant, and the head of every rule of T

- is a literal that does not contain explainable symbols in argument positions, or
- contains no explainable symbols.

It is clear that CBW is definite.

The process of "literal completion" described in [McCain and Turner, 1997] for the propositional case and generalized in [Lifschitz, 1997], turns any definite causal theory into an equivalent set of sentences of classical logic, called its *completion sentences*. If the heads and the bodies of all rules of the theory are first-order formulas, as in CBW, then the completion sentences are first-order also. Applying literal completion is all that needs to be done to eliminate the second-order quantifiers from the sentence representing CBW.

Literal completion is similar to the process of completion familiar from logic programming ([Clark, 1978], [Lloyd and Topor, 1984]). Its main new feature is that the "completed definition" of an explainable predicate constant P consists of *two* equivalences, one "positive" and one "negative." Positive completion sentences are obtained from the rules whose heads are positive literals in exactly the same way as in [Lloyd and Topor, 1984]. Negative completion sentences are generated in a similar way from the rules whose heads are negative literals. In addition, for every rule $F \leftarrow G$ whose head F does not contain explainable symbols, there is a corresponding completion sentence, which is simply the universal closure of $G \supset F$.

In application to CBW, the literal completion procedure generates

(i) the positive completion sentence for *Holds*,

(ii) the negative completion sentence for *Holds*,

(iii) the positive completion sentence for *Inertial*,

(iv) the negative completion sentence for *Inertial*,

(v) the completion sentences corresponding to formulas (11)–(15).

We will consider all these sentences beginning with the easiest, those in group (v). The completion sentence for a rule of form (4) is the universal closure of

$$\neg F \supset \mathit{False},$$

which is equivalent to the universal closure of F. Consequently, sentences (v) are equivalent to the universal closures of (11)–(15).

Sentence (iii) is generated by completing the first two of rules (8). It is equivalent to the universal closure of (16). Sentence (iv) is generated in the same way from the third of rules (8):

$$\forall f(\neg \mathit{Inertial}(f) \equiv \neg \mathit{Inertial}(f)).$$

It is logically valid, and thus can be dropped from the conjunction of (i)–(v).

Sentence (i) is generated from rule (6), the first of rules (9), and the first and third of rules (10). It is equivalent to the universal closure of

$$\begin{aligned}
&\mathit{Holds}(f,s) \equiv \\
&\quad \exists xys_1[f = \mathit{On}(x,y) \wedge s = \mathit{Result}(\mathit{Move}(x,y),s_1) \\
&\qquad \wedge \mathit{Poss}(\mathit{Move}(x,y),s_1)] \\
&\quad \vee \exists as_1[s = \mathit{Result}(a,s_1) \wedge \mathit{Inertial}(f) \\
&\qquad \wedge \mathit{Poss}(a,s_1) \wedge \mathit{Holds}(f,s_1) \wedge \mathit{Holds}(f,s)] \\
&\quad \vee [\mathit{Holds}(f,s) \wedge \neg \mathit{Inertial}(f)] \\
&\quad \vee [\mathit{Holds}(f,s) \wedge \neg \exists as_1(\mathit{Poss}(a,s_1) \\
&\qquad \wedge s = \mathit{Result}(a,s_1))].
\end{aligned} \tag{35}$$

Sentence (ii) is generated from rules (7), the second of rules (9), and the second and fourth of rules (10). It is equivalent to the universal closure of

$$\begin{aligned}
&\neg \mathit{Holds}(f,s) \equiv \\
&\quad \exists xyz[f = \mathit{On}(x,z) \wedge \mathit{Holds}(\mathit{On}(x,y),s) \wedge y \neq z] \\
&\quad \vee \exists xy[f = \mathit{OnT}(x) \wedge \mathit{Holds}(\mathit{On}(x,y),s)] \\
&\quad \vee \exists xy[f = \mathit{On}(x,y) \wedge \mathit{Holds}(\mathit{OnT}(x),s)] \\
&\quad \vee \exists as_1[s = \mathit{Result}(a,s_1) \wedge \mathit{Inertial}(f) \\
&\qquad \wedge \mathit{Poss}(a,s_1) \wedge \neg \mathit{Holds}(f,s_1) \wedge \neg \mathit{Holds}(f,s)] \\
&\quad \vee [\neg \mathit{Holds}(f,s) \wedge \neg \mathit{Inertial}(f)] \\
&\quad \vee [\neg \mathit{Holds}(f,s) \wedge \neg \exists as_1(\mathit{Poss}(a,s_1) \\
&\qquad \wedge s = \mathit{Result}(a,s_1))].
\end{aligned} \tag{36}$$

Thus *CBW* is equivalent to the conjunction of the universal closures of formulas (11)–(16), (35) and (36).

B.2 Proof of Proposition 1

Proposition 1. *CBW is equivalent to the conjunction of the universal closures of formulas (1), (2) and (11)–(19).*

Each of the equivalences (35) and (36) can be written as the conjunction of two implications—left-to-right and right-to-left. We will denote these implications by $(35)^{\supset}$, $(35)^{\subset}$, $(36)^{\supset}$ and $(36)^{\subset}$.

By $\widetilde{\forall} F$ we denote the universal closure of a formula F. We say that formulas F and G are *universally equivalent* if $\widetilde{\forall} F$ is logically equivalent to $\widetilde{\forall} G$. We say that formulas F and G are *universally equivalent in the presence of* a formula H if $\widetilde{\forall} F \wedge \widetilde{\forall} H$ is logically equivalent to $\widetilde{\forall} G \wedge \widetilde{\forall} H$, and similarly for several formulas $H_1, H_2, \ldots$.

In view of the assertion stated at the end of Section B.1, Proposition 1 will be proved if we establish the following lemmas:

Lemma 1. *$(35)^{\subset}$ is universally equivalent to (1).*

Lemma 2. *$(36)^{\subset}$ is universally equivalent to (2).*

Lemma 3. *In the presence of (15), formula $(35)^{\supset}$ is universally equivalent to (17).*

Lemma 4. *In the presence of (15), formula $(36)^{\supset}$ is universally equivalent to (18).*

Proof of Lemma 1. The antecedent of $(35)^{\subset}$ is the disjunction of four formulas, so that $(35)^{\subset}$ can be written as the conjunction of four implications. The first of them is

$$\begin{aligned}
&\exists xys_1[f = \mathit{On}(x,y) \wedge s = \mathit{Result}(\mathit{Move}(x,y),s_1) \\
&\qquad\qquad \wedge \mathit{Poss}(\mathit{Move}(x,y),s_1)] \\
&\qquad\qquad\qquad \supset \mathit{Holds}(f,s),
\end{aligned}$$

which is universally equivalent to (1). The other three implications are tautological.

The proof of Lemma 2 is similar.

Proof of Lemma 3. $(35)^{\supset}$ is equivalent to

$$\begin{aligned}
&\mathit{Holds}(f,s) \\
&\supset \exists xys_1[f = \mathit{On}(x,y) \wedge s = \mathit{Result}(\mathit{Move}(x,y),s_1) \\
&\qquad \wedge \mathit{Poss}(\mathit{Move}(x,y),s_1)] \\
&\vee \exists as_1[s = \mathit{Result}(a,s_1) \wedge \mathit{Inertial}(f) \\
&\qquad \wedge \mathit{Poss}(a,s_1) \wedge \mathit{Holds}(f,s_1)] \\
&\vee \neg \mathit{Inertial}(f) \\
&\vee \neg \exists as_1(\mathit{Poss}(a,s_1) \wedge s = \mathit{Result}(a,s_1))
\end{aligned}$$

which is universally equivalent to

$$\begin{aligned}
&\mathit{Holds}(f,s) \wedge \mathit{Inertial}(f) \\
&\quad \wedge \mathit{Poss}(a,s_1) \wedge s = \mathit{Result}(a,s_1) \\
&\qquad \supset \exists xys_1[f = \mathit{On}(x,y) \\
&\qquad\qquad \wedge s = \mathit{Result}(\mathit{Move}(x,y),s_1) \\
&\qquad\qquad \wedge \mathit{Poss}(\mathit{Move}(x,y),s_1)] \\
&\qquad \vee \exists as_1[s = \mathit{Result}(a,s_1) \wedge \mathit{Poss}(a,s_1) \\
&\qquad\qquad \wedge \mathit{Holds}(f,s_1)].
\end{aligned}$$

In the presence of (15), this is universally equivalent to

$$\begin{aligned}&Holds(f,s) \wedge Inertial(f)\\&\quad\wedge Poss(a,s_1) \wedge s = Result(a,s_1)\\&\qquad\supset \exists xy[f = On(x,y) \wedge a = Move(x,y)\\&\qquad\qquad\wedge Poss(Move(x,y),s_1)]\\&\qquad\vee[Poss(a,s_1) \wedge Holds(f,s_1)]\end{aligned}$$

and consequently to

$$\begin{aligned}&Holds(f,s) \wedge Inertial(f)\\&\quad\wedge Poss(a,s_1) \wedge s = Result(a,s_1)\\&\qquad\supset \exists xy[f = On(x,y) \wedge a = Move(x,y)]\\&\qquad\quad\vee Holds(f,s_1).\end{aligned}$$

This formula is universally equivalent to

$$\begin{aligned}&Holds(f,Result(a,s_1)) \wedge Inertial(f) \wedge Poss(a,s_1)\\&\supset \exists xy[f = On(x,y) \wedge a = Move(x,y)] \vee Holds(f,s_1)\end{aligned}$$

and consequently to (17).

Proof of Lemma 4. $(36)^{\supset}$ is equivalent to

$$\begin{aligned}&\neg Holds(f,s)\\&\supset \exists xyz[f = On(x,z) \wedge Holds(On(x,y),s) \wedge y \neq z]\\&\quad\vee\exists xy[f = OnT(x) \wedge Holds(On(x,y),s)]\\&\quad\vee\exists xy[f = On(x,y) \wedge Holds(OnT(x),s)]\\&\quad\vee\exists as_1[s = Result(a,s_1) \wedge Inertial(f)\\&\qquad\wedge Poss(a,s_1) \wedge \neg Holds(f,s_1)]\\&\quad\vee\neg Inertial(f)\\&\quad\vee\neg\exists as_1(Poss(a,s_1) \wedge s = Result(a,s_1))\end{aligned}$$

which is universally equivalent to

$$\begin{aligned}&\neg Holds(f,s) \wedge Inertial(f) \wedge Poss(a,s_1)\\&\wedge s = Result(a,s_1)\\&\quad\supset \exists xyz[f = On(x,z) \wedge Holds(On(x,y),s) \wedge y \neq z]\\&\quad\vee\exists xy[f = OnT(x) \wedge Holds(On(x,y),s)]\\&\quad\vee\exists xy[f = On(x,y) \wedge Holds(OnT(x),s)]\\&\quad\vee\exists as_1[s = Result(a,s_1) \wedge Poss(a,s_1)\\&\qquad\wedge\neg Holds(f,s_1)].\end{aligned}$$

In the presence of (15), this is universally equivalent to

$$\begin{aligned}&\neg Holds(f,s) \wedge Inertial(f) \wedge Poss(a,s_1)\\&\wedge s = Result(a,s_1)\\&\quad\supset \exists xyz[f = On(x,z) \wedge Holds(On(x,y),s) \wedge y \neq z]\\&\quad\vee\exists xy[f = OnT(x) \wedge Holds(On(x,y),s)]\\&\quad\vee\exists xy[f = On(x,y) \wedge Holds(OnT(x),s)]\\&\quad\vee[Poss(a,s_1) \wedge \neg Holds(f,s_1)]\end{aligned}$$

and consequently to

$$\begin{aligned}&Holds(f,s_1) \wedge \neg Holds(f,s)\\&\wedge Inertial(f) \wedge Poss(a,s_1) \wedge s = Result(a,s_1))\\&\quad\supset \exists xyz[f = On(x,z) \wedge Holds(On(x,y),s) \wedge y \neq z]\\&\quad\vee\exists xy[f = OnT(x) \wedge Holds(On(x,y),s)]\\&\quad\vee\exists xy[f = On(x,y) \wedge Holds(OnT(x),s)].\end{aligned}$$

This is universally equivalent to (18).

B.3 Proof of Corollary to Proposition 1

Corollary. *CBW is equivalent to the conjunction of the universal closures of formulas (1), (2), (11)–(17) and (19).*

It is sufficient to establish the following lemma:

Lemma 5. *The universal closures of (1), (2) and (11)–(17) entail the formulas*

$$\begin{aligned}&Poss(a,s) \wedge Holds(On(x,z),s) \wedge y \neq z\\&\supset [Holds(On(x,y),Result(a,s)) \equiv a = Move(x,y))],\end{aligned} \tag{37}$$

$$\begin{aligned}&Poss(a,s) \wedge Holds(OnT(x),s)\\&\supset [Holds(On(x,y),Result(a,s)) \equiv a = Move(x,y))]\end{aligned} \tag{38}$$

and

$$\begin{aligned}&Poss(a,s) \wedge Holds(On(x,y),s)\\&\qquad\supset \neg Holds(OnT(x),Result(a,s)).\end{aligned} \tag{39}$$

Indeed, the only difference between the statements of Propositions 1 and of the corollary is that (18) is replaced in the latter by (19). By Lemma 5, (18) can be equivalently rewritten as

$$\begin{aligned}&Inertial(f) \wedge Poss(a,s)\\&\wedge Holds(f,s) \wedge \neg Holds(f,Result(a,s))\\&\quad\supset \exists xyz[f = On(x,z) \wedge a = Move(x,y) \wedge y \neq z]\\&\quad\vee\exists xy[f = OnT(x) \wedge a = Move(x,y)]\\&\quad\vee False\end{aligned}$$

which is equivalent to (19).

Proof of Lemma 5. Formula (37) can be written as the conjunction of two, with the equivalence replaced by the implication left-to-right and by the implication right-to-left. We will denote these formulas by $(37)^{\supset}$ and $(37)^{\subset}$. Similarly, (38) can be divided into two parts $(38)^{\supset}$ and $(38)^{\subset}$.

Both $(37)^{\subset}$ and $(38)^{\subset}$ are entailed by (1). From (16) and (17),

$$\begin{aligned}&Poss(a,s) \wedge \neg Holds(On(x,y),s)\\&\wedge Holds(On(x,y),Result(a,s))\\&\quad\supset \exists x_1y_1[On(x,y) = On(x_1,y_1) \wedge a = Move(x_1,y_1)]\end{aligned}$$

and

$$\begin{aligned}&Poss(a,s) \wedge \neg Holds(OnT(x),s)\\&\wedge Holds(OnT(x),Result(a,s))\\&\quad\supset \exists x_1y_1[OnT(x) = On(x_1,y_1) \wedge a = Move(x_1,y_1)].\end{aligned}$$

Then, by (13),

$$\begin{aligned}&Poss(a,s) \wedge \neg Holds(On(x,y),s)\\&\quad\wedge Holds(On(x,y),Result(a,s))\\&\qquad\supset a = Move(x,y))\end{aligned} \tag{40}$$

and

$$\neg[Poss(a,s) \wedge \neg Holds(OnT(x),s) \\ \wedge Holds(OnT(x), Result(a,s))]. \quad (41)$$

Both $(37)^{\supset}$ and $(38)^{\supset}$ are immediate from (2) and (40); (39) is equivalent to (41).

B.4 Proof of Proposition 2

Proposition 2. *The conjunction of LBW with the universal closures of (11)–(15), (23) and (25) is equivalent to the conjunction of CBW with the universal closures of (26) and (27).*

The circumscription LBW is characterized by the following lemma, proved below:

Lemma 6. *LBW is equivalent to the conjunction of the universal closures of (16), (26) and (27).*

On the other hand, by Proposition 1, CBW is equivalent to the conjunction of the universal closures of formulas (1), (2) and (11)–(18). In view of these two facts, we only need to show that, in the presence of the "common axioms" (11)–(16), (26) and (27), the conjunction of (23) and (25) is universally equivalent to the conjunction of (1), (2), (17) and (18). It is easy to check that in the presence of the common axioms

- the first of formulas (23) is universally equivalent to (1),
- the second of formulas (23) is universally equivalent to (2),
- the first of formulas (25) is universally equivalent to (17), and
- the second of formulas (25) is universally equivalent to (18).

Proof of Lemma 6. Recall that LBW is the parallel circumscription of $Caused^{+}$, $Caused^{-}$ and $Inertial$ in the conjunction of the universal closures of formulas (20)–(22). All occurrences of these predicate constants in (20)–(22) are positive. According to Proposition 7.1.1 from [Lifschitz, 1994], it follows that LBW is equivalent to the conjunction of the circumscriptions of each of the three constants in the same sentence. Each of these circumscriptions can be computed using the completion method (see [Lifschitz, 1994], Section 3.1), and the result of the computation is represented by formulas (16), (26) and (27).

B.5 Proof of Proposition 3

Proposition 3. *CBW' is equivalent to the conjunction of the universal closures of formulas (1), (2), (11)–(15), (18), and (30)–(32).*

As discussed in Section B.1, literal completion turns CBW into the equivalent set of sentences (i)–(v). Let us consider how this result is affected by the presence of the additional postulates (28)–(30).

Group (v) includes now the universal closures of formulas (11)–(15) and (30). Sentence (iii) includes a new disjunctive term corresponding to rule (29); it is equivalent to (31). Sentence (iv) is the same logically valid formula as before. Sentence (i) includes a new disjunctive term corresponding to rule (28). It is equivalent to the following modification of (35):

$$\begin{aligned} & Holds(f,s) \equiv \\ & \quad \exists xys_1[f = On(x,y) \wedge s = Result(Move(x,y),s_1) \\ & \qquad \wedge Poss(Move(x,y),s_1)] \\ & \quad \vee \exists xys_1[f = Damaged(x) \\ & \qquad \wedge s = Result(Move(x,y),s_1) \\ & \qquad \wedge Poss(Move(x,y),s_1) \\ & \qquad \wedge Holds(Damaged(x), Result(Move(x,y),s_1))] \\ & \quad \vee \exists as_1[s = Result(a,s_1) \wedge Inertial(f) \\ & \qquad \wedge Poss(a,s_1) \wedge Holds(f,s_1) \wedge Holds(f,s)] \\ & \quad \vee [Holds(f,s) \wedge \neg Inertial(f)] \\ & \quad \vee [Holds(f,s) \wedge \neg\exists as_1(Poss(a,s_1) \\ & \qquad \wedge s = Result(a,s_1))]. \end{aligned} \quad (42)$$

Finally, sentence (ii) is the same before, and it is equivalent to the universal closure of (36).

Consequently CBW' is equivalent to the conjunction of the universal closures of formulas (11)–(15), (30), (31), (36) and (42).

The assertion of Proposition 3 follows from this fact in view of Lemmas 2 and 4 and the following two assertions, characterizing the two halves of the equivalence (42):

Lemma 7. *$(42)^{\subset}$ is universally equivalent to (1).*

Lemma 8. *In the presence of (15), formula $(42)^{\supset}$ is universally equivalent to (32).*

Proof of Lemma 7. Similar to the proof of Lemma 1; $(42)^{\subset}$ can be written as the conjunction of five implications, and the implication containing $Damaged$ is logically valid.

Proof of Lemma 8. The proof is similar to the proof of Lemma 3. After the first two simplification steps, we get:

$$
\begin{array}{l}
Holds(f,s) \wedge Inertial(f) \\
\wedge Poss(a,s_1) \wedge s = Result(a,s_1) \\
\supset \exists x y s_1 [f = On(x,y) \\
\quad \wedge s = Result(Move(x,y),s_1) \\
\quad \wedge Poss(Move(x,y),s_1)] \\
\vee \exists x y s_1 [f = Damaged(x) \\
\quad \wedge s = Result(Move(x,y),s_1) \\
\quad \wedge Poss(Move(x,y),s_1) \\
\quad \wedge Holds(Damaged(x), Result(Move(x,y),s_1))] \\
\vee \exists a s_1 [s = Result(a,s_1) \wedge Poss(a,s_1) \\
\quad \wedge Holds(f,s_1)].
\end{array}
$$

In the presence of (15), this is universally equivalent to

$$
\begin{array}{l}
Holds(f,s) \wedge Inertial(f) \\
\wedge Poss(a,s_1) \wedge s = Result(a,s_1) \\
\supset \exists x y [f = On(x,y) \wedge a = Move(x,y) \\
\quad \wedge Poss(Move(x,y),s_1)] \\
\vee \exists x y [f = Damaged(x) \\
\quad \wedge a = Move(x,y) \\
\quad \wedge Poss(Move(x,y),s_1) \\
\quad \wedge Holds(Damaged(x), Result(Move(x,y),s_1))] \\
\vee [Poss(a,s_1) \wedge Holds(f,s_1)]
\end{array}
$$

and consequently to

$$
\begin{array}{l}
Holds(f,s) \wedge Inertial(f) \\
\wedge Poss(a,s_1) \wedge s = Result(a,s_1) \\
\quad \supset \exists x y [f = On(x,y) \wedge a = Move(x,y)] \\
\quad \vee \exists x y [f = Damaged(x) \wedge a = Move(x,y)] \\
\quad \vee Holds(f,s_1).
\end{array}
$$

This formula is universally equivalent to

$$
\begin{array}{l}
Holds(f, Result(a,s_1)) \wedge Inertial(f) \wedge Poss(a,s_1) \\
\supset \exists x y [f = On(x,y) \wedge a = Move(x,y)] \\
\quad \vee \exists x y [f = Damaged(x) \wedge a = Move(x,y)] \\
\quad \vee Holds(f,s_1)
\end{array}
$$

and consequently to (32).

References

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

[Clark, 1978] Keith Clark. Negation as failure. In Herve Gallaire and Jack Minker, editors, *Logic and Data Bases*, pages 293–322. Plenum Press, New York, 1978.

[Geffner, 1990] Hector Geffner. Causal theories for nonmonotonic reasoning. In *Proc. AAAI-90*, pages 524–530, 1990.

[Giunchiglia and Lifschitz, 1998] Enrico Giunchiglia and Vladimir Lifschitz. An action language based on causal explanation: Preliminary report. In *Proc. AAAI-98*, 1998. To appear.

[Kartha, 1993] Neelakantan Kartha. Soundness and completeness theorems for three formalizations of action. In *Proc. IJCAI-93*, pages 724–729, 1993.

[Lifschitz, 1991] Vladimir Lifschitz. Towards a metatheory of action. In James Allen, Richard Fikes, and Erik Sandewall, editors, *Proc. Second Int'l Conf. on Principles of Knowledge Representation and Reasoning*, pages 376–386, 1991.

[Lifschitz, 1994] Vladimir Lifschitz. Circumscription. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *The Handbook of Logic in AI and Logic Programming*, volume 3, pages 298–352. Oxford University Press, 1994.

[Lifschitz, 1997] Vladimir Lifschitz. On the logic of causal explanation. *Artificial Intelligence*, 96:451–465, 1997.

[Lin and Reiter, 1994] Fangzhen Lin and Raymond Reiter. State constraints revisited. *Journal of Logic and Computation*, 4:655–678, 1994.

[Lin, 1995] Fangzhen Lin. Embracing causality in specifying the indirect effects of actions. In *Proc. IJCAI-95*, pages 1985–1991, 1995.

[Lloyd and Topor, 1984] John Lloyd and Rodney Topor. Making Prolog more expressive. *Journal of Logic Programming*, 3:225–240, 1984.

[McCain and Turner, 1995] Norman McCain and Hudson Turner. A causal theory of ramifications and qualifications. In *Proc. IJCAI-95*, pages 1978–1984, 1995.

[McCain and Turner, 1997] Norman McCain and Hudson Turner. Causal theories of action and change. In *Proc. AAAI-97*, pages 460–465, 1997.

[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, 1990] John McCarthy. *Formalizing Common Sense: Papers by John McCarthy.* Ablex, Norwood, NJ, 1990.

[Reiter, 1991] Raymond Reiter. The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression.

In Vladimir Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*, pages 359–380. Academic Press, 1991.

[Schubert, 1990] Lenhart Schubert. Monotonic solution of the frame problem in the situation calculus: an efficient method for worlds with fully specified actions. In H.E. Kyburg, R. Loui, and G. Carlson, editors, *Knowledge Representation and Defeasible Reasoning*, pages 23–67. Kluwer, 1990.

[Turner, 1998] Hudson Turner. A logic of universal causation. *Artificial Intelligence*, 1998. To appear.