Conditionals and Action Logics*

Richmond H. Thomason
Philosophy Department
University of Michigan
Ann Arbor, MI 48109-1003
U.S.A.

rich@thomason.org

Abstract

The logic of conditionals (and, especially, of subjunctive or counterfactual conditionals) is a long-standing problem for the theory of common sense reasoning and philosophical logic. In this paper, I explore commonalities between the logical accounts of the conditional involving "closeness" relations over possible worlds and work in the logic of action and change that attempts to characterize the situation that results from the performance of an action. I will try to show that the latter approaches readily generalize to the case of conditionals (with suitably restricted antecedents), and that they fill a gap in the earlier, more abstract work on conditional logic.

This is an abbreviated version for the CSR-07 workshop.

Stalnaker's and Lewis' conditional logics

Counterfactual conditionals are a prominent and pervasive part of common sense reasoning,[1] but accounting for their meaning has been a long-standing problem in philosophy and logic.

In the late 1960s (Stalnaker 1968; Stalnaker & Thomason 1970) and early 1970s (Lewis 1973), theories of conditionals were developed using the possible worlds approach to modal logic. These theories appealed to "closeness" relations over possible worlds. Stalnaker's semantics in effect posited a well-ordering $\preceq_w$ over worlds with respect to an initial world w, with w as a least element. A conditional $\phi > \psi$ is true at w if and only if ψ is true at the closest world to w in which ϕ is true, if there is any such world; otherwise, $\phi > \psi$ is true at w. To put it another way, a Stalnaker frame includes a selection function s that for each world w and antecedent ϕ selects the world $s(w, \phi)$ that results from changing w so as to make ϕ true. The completeness theorem in (Stalnaker & Thomason 1970) shows how to recover appropriate well-orderings in canonical models of the axioms for conditional logic provided in that paper.

Lewis' semantics is more complicated; the chief difference between it and Stalnaker's is that $(\phi > \psi) \vee (\phi > \neg\psi)$ (Conditional Excluded Middle) is valid in Stalnaker's logic, but is invalid in Lewis'.[2]

Although there has been much discussion in the subsequent literature over the details of the semantics, and especially over the semantic differences between indicative and subjunctive conditionals, logics along these lines remain the standard approach to subjunctive conditionals. Many people (if not a majority) still defend Stalnaker's original claim that there is one semantics for all conditionals, and that the differences between subjunctive and indicative conditionals are pragmatic. But even if these people are right, the differences between indicative (or epistemic) and subjunctive (or causal) conditionals are significant, and play an important role in reasoning.

In this paper, I will work with Stalnaker's semantics and will confine myself to subjunctive or causal conditionals.[3] But a good deal of what I will say could be adapted to other logics of conditionals.

An epistemological problem

Work on causal conditionals prior to WW2 consisted, for the most part, of attempts to provide a philosophical analysis. In a landmark paper (Goodman 1955) dating to 1946, Nelson Goodman presented the problems confronting such an analysis so devastatingly that these attempts were largely abandoned.

I don't have space here to examine Goodman's arguments. They deal mainly with a single example—'If that match had been scratched it would have lighted'—and show that attempts to provide an analysis either bog down or prove to be circular. The problems that Goodman raises are, I believe, genuine, although his methodological assumptions are flawed in some fundamental ways, and he tends to exaggerate the difficulties.

The logical solutions of the 1960s and 1970s end-run Goodman's problem, rather than meeting it head-on. They provide abstract conditions on modal models for conditionals rather than analyses. Actually constructing the Stalnaker

*Thanks to the referees of this paper for helpful comments.

[1](McCarthy & Costello 1999) makes this point well.

[2]Therefore, Lewis uses a partial order over worlds, while Stalnaker uses a total order. The complexity of Lewis' semantic conditions have to do with his treatment of the "limit assumption" and are not relevant to the present project.

[3]Although these conditionals are connected with subjunctive mood in some languages, the connection is tenuous in English. From here on, I will call them "causal conditionals."

selection function for a particular domain would raise problems like Goodman's, but this construction is not needed to define a notion of logical consequence. You could say (and many of the people who adopted these theories did say) that they provide a solution to the logical problem of conditionals, but avoid the epistemological problem of conditionals.

Despite the success of the logical theories, I believe that most of us involved in this episode realized that there was a gap. For instance, a number of papers were published criticizing the logical theories by pointing out that conditionals do not align with the closest world, if "closest" is measured by apparent overall similarity. (On this interpretation, worlds that could not be perceptually distinguished would be very close; worlds that are very different in many important ways would be far apart.) Since small actions can have sweeping effects, this crude notion of closeness will obviously not work for causal conditionals. The obvious answer to this objection is that the closeness that is appropriate for causal conditionals is not based on superficial similarity. But if you try to supplement this with a positive, specific, constructive account of the correct notion of "similarity," you re-encounter Goodman's problems.

You can try (and people have tried) to produce such an account using philosophical methods, but I believe that this project exceeds the capacity of the methods that have been traditionally used in philosophy, and that still prevail in the field.

The literature in AI and CSR

AI has its own (much smaller) literature on conditionals; (Ginsberg 1986; Ortiz, Jr. 1999b; 1999a) are especially noteworthy. I do not have space here to discuss this work; it contains many good ideas, but I believe that it is flawed by a failure to carefully distinguish between causal and epistemological conditionals. And without making this distinction, it is hard to see the relevance of work on action and change to conditionals. As far as I know, the closest work in the AI literature to the approach I adopt here is (Pearl 2000). Pearl uses a very different, statistical framework, and the similarities lie mainly in the basic motivating ideas.

Using an action logic to model simple conditionals

I want to argue here that the AI literature on formalisms for action and change provides a solution to the epistemological problem of causal conditionals by allowing selection functions for these conditionals to be constructed for nontrivial domains. Work in this area over the last 20 years has in fact addressed the major methodological problems raised by Goodman. The methodology that has emerged in AI is similar in some ways to philosophical analysis, but by concentrating on reasoning about action and change it has turned a notorious philosophical stumbling block into an incremental research program. Most important for this success, I believe, are the use of special-purpose logics designed to overcome formalization difficulties, and of limited domains and benchmark problems as testing grounds for ideas.

Language and Models

The remainder of this paper illustrates how an action-and-change logic can provide models for conditionals.

I will use the "causal rule" approach of (Giunchiglia *et al.* 2004), which has the advantage of being relatively simple, while at the same time it incorporates a causal update mechanism that is general enough to extend to conditionals. Also, axiomatizations of moderately large domains have been undertaken in connection with this formalism, and this paper—brief as it is—should make a convincing case that these axiomatizations could be fairly easily adapted to provide semantics for conditionals. However, other approaches, such as the Event Calculus (Kowalski & Sergot 1986) or GoLog (Reiter 2001) could have been used as well, and these would have made other domain axiomatizations available.[4]

Recall that, on Stalnaker's approach, a model for propositional conditional logic consists of (1) A set W of possible worlds, (2) an assignment V of a subset $V(p)$ of W to each atomic formula p, and (3) a function s from formulas and e-worlds to e-worlds. (An e-world is either a member of W or a designated "absurd world" ω. The absurd world is not a world—it is a convenient way of managing inconsistent antecedents.) Our project here is to use the apparatus of this causal logic to define an appropriate s.

I will begin with a special case, in which the antecedents of conditionals are action declarations. I will work with propositional languages that are based on a set of actions and a set of propositional actions. The simplest such language is closed under boolean operations and has a single action as the antecedent of a conditional. This simplification confines the theory to *first-degree* formulas, in which there is no nesting of conditionals.

Specializing Stalnaker's selection function semantics for conditionals to this case requires us to take time into consideration and to accommodate the fact that counterfactual transitions are driven by actions rather than by arbitrary formulas. First, we reinterpret the worlds of our models temporally, as world-time pairs. For our simple languages, which have no tense operators, this requires no formal changes; in fact, we will simply construe the elements of W as world-time pairs, even though we continue to speak of them as worlds. (More often, I will simply call them points.") Second, the selection function will now take worlds and actions into worlds. Such a function, of course, is exactly what is delivered by a deterministic action logic.

Definition 1. The language $\mathcal{AC}_0$.

Let $\mathcal{A}$ and $\mathcal{P}$ be disjoint sets. Then (1) any member of $\mathcal{P}$ is a formula of $\mathcal{AC}_0$; (2) $\perp$ is a formula of $\mathcal{AC}_0$; (3) if ϕ and ψ are formulas of $\mathcal{AC}_0$ then $\phi \rightarrow \psi$ is a formula of $\mathcal{AC}_0$; and (4) if $a \in \mathcal{A}$ and ϕ is a formula of $\mathcal{AC}_0$ then $a > \phi$ is a formula of $\mathcal{AC}_0$.

The full set of boolean operators can be defined in the usual way; for instance, $\neg\phi$ is $\phi \rightarrow \perp$. $\top$ is $\neg\perp$.

[4] For instance, an adaptation of the Event Calculus similar to the one described here of the Causal Calculus, in connection with Shanahan's solution to the Egg Cracking Problem (Shanahan 2001) would deliver a semantics for conditionals in that domain.

Definition 2. Worlds, extended worlds, frames.
Where ω is a designated nonelement of W, W^+ is $W \cup \{\omega\}$. A *Stalnaker frame* for $\mathcal{AC}_0$ is a pair $\langle W, s\rangle$, where W is a nonempty set and s is a function from $W \times \mathcal{A}$ to W^+.

Usually in modal logic, we postulate a frame as part of a model and then go on to investigate validity. Here the problem is to *construct* the frame from a theory that lends itself to the formalization of challenging domains.

In reconstructing a semantics for conditional logic, we are committed to a modal syntax in which worlds are not named or built into atomic formulas. This difference between conditional logic and the language and formalization style of (Giunchiglia *et al.* 2004) will require some changes to the presentation of the Causal Calculus, in which some elements of the theory are shifted from the syntax to the semantics.[5]

We begin with an adaptation of the fundamental elements of the Causal Logic approach: causal theories, interpretations, and models.

Causal rules are like the rules of (Giunchiglia *et al.* 2004), but are indexed with actions.

Definition 3. Causal rule, causal theory.
Where ϕ and ψ are propositional formulas of $\mathcal{AC}_0$ and a is an action, $\psi \Leftarrow [a]\phi$ is an $\mathcal{AC}_0$ *causal rule*. An $\mathcal{AC}_0$ *causal theory* is a set T of $\mathcal{AC}_0$ causal rules.

In (Giunchiglia *et al.* 2004), interpretations are temporally global, in the sense that they interpret a language with formulas indexed to many timepoints. Our modal adaptation is confined to transitions—temporally adjacent pairs of points—so we can split the interpretations of (Giunchiglia *et al.* 2004) into two interpretations, one for the initial point of the transition and one for the point that ensues when an action a is performed.

Definition 4. P-interpretation of a language.
A propositional *interpretation* or *p-interpretation* I of $\mathcal{AC}_0$ is a pair $\langle I_1, I_2\rangle$ of subsets of $\mathcal{P}$.

Definition 5. Satisfaction by a P-interpretation.
Where ϕ is a propositional formula of $\mathcal{AC}_0$ $I_1 \models \phi$ and $I_2 \models \phi$ are defined in the familiar way.

Definition 6. $\Delta(T, I, a)$.
Let T be a causal theory, let $I = \langle I_1, I_2\rangle$ be a p-interpretation, and let $a \in \mathcal{A}$. Then $\Delta(T, I, a) =$
$\{\psi \ / \ I_1 \models \phi$ for some causal rule $\psi \Leftarrow [a]\phi \in T\}$.

Definition 7. Model of a causal theory and action a.
Let T be a causal theory, let I be a p-interpretation of $\mathcal{AC}_0$, and let $a \in \mathcal{A}$. I is a *model* of T for a iff $I_2 \models \phi$ for all $\phi \in \Delta(T, I, a)$.

Definition 8. Causal model of a causal theory and action a.
Let T be a causal theory and $a \in \mathcal{A}$. A *causal model* or *c-model* of a T for a in $\mathcal{AC}_0$ is a p-interpretation I of $\mathcal{AC}_0$ that is a unique model of T.

The crucial idea of satisfaction in a c-model, and how c-models deliver a nonmonotonic logical consequence relation, are well explained in (Giunchiglia *et al.* 2004) (where c-models simply called "models").

[5] Alternatively, we could, of course, work with a hybrid formalization of conditional logic.

A simple example

We will work with the match-lighting domain of (Goodman 1955). There are two matches, which can be wet or dry, lit or unlit, and struck or unstruck; oxygen can be present or not. There are two striking actions, one for each match.

In Causal Logic, constraints on the initial conditions are stated contrapositively; a causal axiom of the form $\bot \Leftarrow [a]\neg\phi$ requires ϕ to hold in the initial state. To axiomatize a complete world, we begin with an intended initial interpretation I_1 and let *World*$[a](I_1)$ be $\{\bot \Leftarrow [a]\neg\eta \ / \eta \in I_1\} \cup \{\bot \Leftarrow [a]\eta \ / \eta \notin I_1\}$. These axioms guarantee that the initial world will match I_1.

The following axioms aim at an initial state in which oxygen is present and both matches are dry, unlit, and unstruck. They give the dynamics for the action of striking the first match. The axioms assume that a match that has been struck will not light if struck again; this is an oversimplification, but I don't believe we can do better in the context of a deterministic framework without introducing a hidden variable.

See p. 6 for Figure 1

The intended model is $I = \langle I_1, I_2\rangle$, where:

I_1 = $\{\mathtt{oxygen}, \mathtt{dry}_1, \mathtt{dry}_2, \neg\mathtt{lit}_1, \neg\mathtt{lit}_2, \neg\mathtt{struck}_1, \neg\mathtt{struck}_2\}$

I_2 = $\{\mathtt{oxygen}, \mathtt{dry}_1, \mathtt{dry}_2, liti, \neg\mathtt{lit}_2, \mathtt{struck}_1, \neg\mathtt{struck}_2\}$

In this example,
$\Delta(T, I, \mathtt{strike}_1) = \{\mathtt{oxygen}, \mathtt{dry}_1, \mathtt{dry}_2, \neg\mathtt{lit}_2, \mathtt{lit}_1, \mathtt{struck}_1, \neg\mathtt{struck}_2\}$.
Therefore, $I2 \models \Delta(T, I, \mathtt{strike}_1)$, so that I models T.

We now show that I is unique. Let $I' = \langle I'_1, I'_2\rangle$, and suppose that $I \models T$. Then:
$I'_1 \models \{\mathtt{dry}_1, \mathtt{dry}_2, \mathtt{oxygen}, \neg\mathtt{struck}_1, \neg\mathtt{struck}_2, \neg\mathtt{lit}_1, \neg\mathtt{lit}_2\}$.
Therefore, $\Delta(T, I', \mathtt{strike}_1) = \{\mathtt{oxygen}, \mathtt{dry}_1, \mathtt{dry}_2, \mathtt{lit}_1, \neg\mathtt{lit}_2, \mathtt{struck}_1, \neg\mathtt{struck}_2\}$. So I=I', and I is a model of T.

The axiomatization of the domain is easily completed with analogous axioms for $\mathtt{strike}_2$. We now show how to generate a Stalnaker frame for $\mathcal{AC}_0$ from the causal axioms.

To simplify matters, we will identify worlds with consistent, complete sets of literals—i.e., with consistent sets of literals that contain each literal or its negation. In particular, we are interested in the following worlds.

w_0 : $\{\mathtt{dry}_1, \mathtt{dry}_2, \mathtt{oxygen}, \neg\mathtt{struck}_1, \neg\mathtt{struck}_2, \neg\mathtt{lit}_1, \neg\mathtt{lit}_2\}$

w_1 : $\{\mathtt{dry}_1, \mathtt{dry}_2, \mathtt{oxygen}, \mathtt{struck}_1, \neg\mathtt{struck}_2, \mathtt{lit}_1, \neg\mathtt{lit}_2\}$

w_2 : $\{\mathtt{dry}_1, \mathtt{dry}_2, \mathtt{oxygen}, \neg\mathtt{struck}_1, \mathtt{struck}_2, \neg\mathtt{lit}_1, \mathtt{lit}_2\}$

w_3 : $\{\mathtt{dry}_1, \mathtt{dry}_2, \mathtt{oxygen}, \mathtt{struck}_1, \mathtt{struck}_2, \mathtt{lit}_1, \mathtt{lit}_2\}$

The axioms induce the following Stalnaker selection function over these worlds.

See p. 6 for Figure 2

The selection function that emerges from this treatment deals in a principled way with the main problems raised in (Goodman 1955). It provides a well-motivated formalization that corresponds well to intuitive judgments about what would happen if the match were struck in a domain that includes the variables with which Goodman was concerned, and does so in a way that can be extended to wider domains.

Of course, the formalization is not entirely problem-free. (1) Other factors that are not included in the model may prevent the striking action from succeeding. (2) When one match is lit, it may set off the other match. (3) If a match fails to strike at first, it may light when struck again. (4) When a match is lit, it stays lit briefly—the fluent $\texttt{lit}_1$ is inertial for only for a short, indefinite period of time.

All of these problems are considered in the literature on logics of action and change. Problem (1) is the qualification problem. As a practical problem in axiomatization, it can be dealt with using a nonmonotonic logic, so that new qualifications can be treated as additional axioms, without withdrawing previous axioms. Problem (2) can be dealt with by introducing a new fluent tracking the closeness of the matches when struck. Problem (3), which was already mentioned, may call for a nondeterministic model. Problem (4) seems to call for fairly sweeping changes, perhaps introducing introducing ideas from qualitative (or even quantitative) physics into the model.

Of these problems, Goodman mentions only the first. He seems to regard it as obviously insoluble, but gives no arguments about why it can't be dealt with systematically.

Causal theories are used in the account of conditionals, and these theories involve rules like $\texttt{dry}_1 \Leftarrow [\texttt{strike}_1]\texttt{dry}_1$, which look like conditionals. This does not introduce a circularity into the account—causal rules are very different from causal conditionals. The success of this approach depends on how well motivated Causal Logic is, and on whether it can be successfully used to model large causal domains.

Modeling more complex conditionals with action logic

The language $\mathcal{AC}_0$ is very impoverished, even though it covers Goodman's showcase example. In particular, it doesn't allow us to put formulas into the antecedents of conditionals, or to formulate examples of the inference patterns that characterize conditional logics. For instance, we can't formulate instances of *strengthening the antecedent*: from $p > r$ to infer $(p \wedge q) > r$.

We now turn to a logic $\mathcal{AC}_1$, which—although it is still a fragment of conditional logic—is considerably more expressive than $\mathcal{AC}_0$.

Definition 9. The language $\mathcal{AC}_1$.

Let $\mathcal{A}$ and $\mathcal{P}$ be disjoint sets. Then (1) any member of $\mathcal{P}$ is a formula of $\mathcal{AC}_1$; (2) $\perp$ is a formula of $\mathcal{AC}_1$; (3) if ϕ and ψ are formulas of $\mathcal{AC}_1$ then $\phi \rightarrow \psi$ is a formula of $\mathcal{AC}_1$; and (4) if $a \in \mathcal{A}$ and $\eta_1, \ldots, \eta_n$ are literals of $\mathcal{AC}_1$ then $\eta_1, \ldots, \eta_n, a > \psi$ are formulas of $\mathcal{AC}_1$.

The antecedents of conditionals in $\mathcal{AC}_1$, which consist in part of a conjunction of literals, require us to consider counterfactual transitions that are induced by declarative hypotheses rather than by actions: hypotheses like (i) 'If this match were wet, the other match would be wet' or even like (ii) 'If this match were wet, it wouldn't light if struck'.

The counterfactual transitions needed for (ii) can be decomposed into two steps: (1) from the initial point w_0 to a point w_1 that is "simultaneous"[6] with the w_0, but that differs from w_0 in satisfying a specified conjunction of literals, and (2) from w_1 to the point w_2 that results from performing a given action in w_1.

The second sort of transition we have already dealt with. And ideas from action and change formalisms can be adapted to the first transition. Although these transitions are not induced by actions, the changes that are constructed in formalisms like Causal Logic are actually produced not by the actions themselves, but by the immediate causal consequences of these actions. In general, they would be recorded in the form of axioms of the form

$$(\textit{Preconds}(\mathrm{a}) \wedge \textit{InitialConds}(a, \eta)) \rightarrow \eta,$$

where $\textit{Preconds}(a)$ formulates the preconditions of a and $\textit{InitialConds}(a, \eta)$ the initial conditions under which η will be directly caused by a performance of a.

The antecedent of a subjunctive conditional explicitly states the "direct causal consequences," so the need for preconditions and initial conditions in these axioms drops out, and, where an antecedent contains a literal η, all we need is an axiom saying that η holds in the ensuing point.[7]

To accommodate these ideas in a version of the Causal Calculus, we need two sorts of causal laws and three-part interpretations. Below, I present only the definitions that need to be amended in passing to $\mathcal{AC}_1$.

Definition 10. Causal rule, causal theory.

Where ϕ and ψ are propositional formulas of $\mathcal{AC}_0$, a is an action, and $\eta_1, \ldots, \eta_n$ are literals, $\psi \Leftarrow [a]\phi$ and $\psi \Leftarrow [\eta_1, \ldots, \eta_n]\phi$ are $\mathcal{AC}_1$ *causal rules*. An $\mathcal{AC}_1$ *causal theory* is a set T of causal rules.

Definition 11. P-interpretation of a language.

An ACi *p-interpretation* I of $\mathcal{AC}_0$ is a triple $\langle I_1, I_2, I_3 \rangle$ of subsets of $\mathcal{P}$.

Definition 12. $\Delta_1(T, I, \{\eta_1, \ldots, \eta_n\})$.

$\Delta_1(T, I, \{\eta_1, \ldots, \eta_n\}) = \{\eta_1, \ldots, \eta_n\} \cup$
$\{\psi \ / \ I_1 \models \phi$ for some causal rule
$\psi \Leftarrow [\eta_1, \ldots, \eta_n]\phi \in T\}$.

Definition 13. $\Delta_2(T, I, a)$.

Where T is a causal theory, $I = \langle I_1, I_2, I_3 \rangle$ is a p-interpretation, and $a \in \mathcal{A}$, $\Delta_2(T, I, a) =$
$\{\psi \ / \ I_2 \models \phi$ for some causal rule $\psi \Leftarrow [a]\phi \in T\}$.

Definition 14. Model of an $\mathcal{AC}_1$ causal theory and action a.

Let T be an $\mathcal{AC}_1$ causal theory, let I be a p-interpretation of $\mathcal{AC}_1$, let $\eta_1, \ldots, \eta_n$ be literals, and let $a \in \mathcal{A}$. I is a *model* of T for $\eta_1, \ldots, \eta_n$ and a iff $I_2 \models \phi$ for all $\phi \in \Delta_1(T, I, a)$ and $I_3 \models \phi$ for all $\phi \in \Delta_2(T, I, a)$.

[6] This can be an abstract simultaneity relation—we do not need to resort to a temporal metric. For a general treatment of conditionals and time, see (Thomason & Gupta 1980).

[7] This idea is similar to Pearl's characterization of counterfactuals using "principled minisurgery" operators $do(X = x)$. See (Pearl 2000, §7.1).

Notice that the "causal effects" declarative transitions are built into Definition 12.

Our intuitions about declarative counterfactual transitions are somewhat less robust that those about action transitions, probably because we have in general a good idea of *how* an action will take place. (Does a condition like 'If that wet match were dry' envisage a state in which it never became wet, or a state in which it was somehow dried after it became wet?) Nevertheless, we do have clear enough intuitions to write axioms, and I believe that the model I have proposed will deliver pretty good results.

To illustrate these ideas, we extend the match domain to include a new fluent `together`, tracking whether the matches are together. There are also two new actions, which spill water on the matches. This will allow us to introduce some ramifications; we suppose that if the matches are together, both will be wet if either would be wet.

Because of space limitations, the extended theory must be sketched. The following is a sample of the required axioms.[8]

See p. 6 for Figure 3

With these axioms, for instance, we satisfy the conditional 'If match 2 were wet and the matches were together, then match 1 would be wet', and 'If match 2 were wet and the matches were together and match 1 were struck, then match 2 would not be lit'.

Concluding remarks

In providing a selection function, this approach fills another gap left open by the bare possible worlds semantics for conditionals, by explaining the truth of "inertial" conditionals, such as 'If match 1 were struck then match 2 would (still) be dry', which is true when match 2 is dry.

However, it provides a selection function only for a fragment of even the language of first-degree conditionals. To complete the first-degree semantics, we would need to have a function taking a set of alternative conjunctions of literals into a single preferred alternative. I don't see at the moment how to generate such preferences in a principled way.[9]

Extending the theory to deal with past counterfactuals like 'If I had struck the match it would have lit' seems more promising. These conditionals seem to be true if the corresponding present counterfactual was true at an appropriate time. 'If I had struck the match it would have lit' is true now if at a previous time, 'If I were to strike the match it would light' is true. We have already given an account of the latter conditionals. So (roughly) we can say that a past counterfactual $a > \phi$ is true in case at (say) the closest state in the past at which the preconditions of a hold, ϕ is true at a point simultaneous with the present along a counterfactual history beginning with the performance of a, and resembling in some way the actual history. The main problem here is figuring out how to make the resemblance precise.

With declarative counterfactual axioms, we are faced in the worst case with an axiom for each set of literals in the language. The resulting explosion of axioms is probably the most worrisome problem with the approach that I have sketched here. But the problem is not specific to the approach I have undertaken here—any attempt to construct a selection function will encounter it, because of the invalidity of weakening the antecedent.

Nevertheless we can reason effectively with counterfactuals in many commonsense domains. I believe this is because counterfactual independence conditions are legitimate in these cases. Hopefully, these independence conditions can be used to keep the reasoning from becoming hopelessly intractable in realistic domains. Causal graphs might be used for this purpose, but I have not yet explored this line of inquiry.

[8] These axioms are schemes: '$\texttt{oxygen} \Leftarrow [L]\texttt{oxygen}$ if $\neg\texttt{oxygen} \notin L$' for instance, stands for a set of axioms, one for each set L of literals such that $\neg\texttt{oxygen} \notin L$'.

[9] Disjunctive antecedents are often used by skeptics who question the meaningfulness of conditionals. Quine asks: "If Bizet and Verdi had been compatriots, would Verdi have been French or Bizet have been Italian?

References

Ginsberg, M. L. 1986. Counterfactuals. *Artificial Intelligence* 30(1):35–79.

Giunchiglia, E.; Lee, J.; Lifschitz, V.; McCain, N.; and Turner, H. 2004. Nonmonotonic causal theories. *Artificial Intelligence* 153(5–6):49–104.

Goodman, N. 1955. *Fact, Fiction and Forecast*. Harvard University Press.

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

Lewis, D. K. 1973. *Counterfactuals*. Cambridge, Massachusetts: Harvard University Press.

McCarthy, J., and Costello, T. 1999. Useful counterfactuals. *Linköping Electronic Articles in Computer and Information Science* 4(12). Available at http://www.ep.liu.se/ea/cis/1999/012/.

Ortiz, Jr., C. L. 1999a. A commonsense language for reasoning about causation and rational action. *Artificial Intelligence* 111(1–2):73–169.

Ortiz, Jr., C. L. 1999b. Explanatory update theory: Applications of counterfactual reasoning to causation. *Artificial Intelligence* 108(1–2):125–178.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, England: Cambridge University Press.

Reiter, R. 2001. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. Cambridge, Massachusetts: The MIT Press.

Shanahan, M. 2001. An attempt to formalize a nontrivial benchmark problem in common sense reasoning. *Artificial Intelligence* 153(5–6):141–165.

Stalnaker, R. C., and Thomason, R. H. 1970. A semantic analysis of conditional logic. *Theoria* 36:23–42.

Stalnaker, R. C. 1968. A theory of conditionals. In Rescher, N., ed., *Studies in Logical Theory*. Oxford: Basil Blackwell Publishers. 98–112.

Thomason, R., and Gupta, A. 1980. A theory of conditionals in the context of branching time. *The Philosophical Review* 80:65–90.

Fluents:	$\texttt{oxygen}, \texttt{lit}_1, \texttt{lit}_2, \texttt{dry}_1, \texttt{dry}_2, \texttt{struck}_1, \texttt{struck}_2$
Actions:	$\texttt{strike}_1, \texttt{strike}_2$
Intended I_1:	$\{\texttt{dry}_1, \texttt{dry}_2, \texttt{oxygen}, \neg\texttt{struck}_1, \neg\texttt{struck}_2, \neg\texttt{lit}_1, \neg\texttt{lit}_2\}$
Initial conditions:	$\mathit{World}[(]I_1)$
Inertial axioms:	$\texttt{oxygen} \Leftarrow [\texttt{strike}_1]\texttt{oxygen}$ $\neg\texttt{oxygen} \Leftarrow [\texttt{strike}_1]\neg\texttt{oxygen}$ $\texttt{dry}_1 \Leftarrow [\texttt{strike}_1]\texttt{dry}_1$ $\neg\texttt{dry}_1 \Leftarrow [\texttt{strike}_1]\neg\texttt{dry}_1$ $\texttt{dry}_2 \Leftarrow [\texttt{strike}_1]\texttt{dry}_2$ $\neg\texttt{dry}_2 \Leftarrow [\texttt{strike}_1]\neg\texttt{dry}_2$ $\mathit{struckii} \Leftarrow [\texttt{strike}_1]\mathit{struckii}$ $\neg\texttt{struck}_2 \Leftarrow [\texttt{strike}_1]\neg\texttt{struck}_2$ $\texttt{lit}_2 \Leftarrow [\texttt{strike}_1]\texttt{lit}_2$ $\neg\texttt{lit}_2 \Leftarrow [\texttt{strike}_1]\neg\texttt{lit}_2$
Change Axioms:	$\texttt{struck}_1 \Leftarrow [\texttt{strike}_1]\neg\texttt{struck}_1$ $\texttt{lit}_1 \Leftarrow [\texttt{strike}_1]\texttt{oxygen}, \texttt{dry}_1, \neg\texttt{struck}_1$

Figure 1: Axioms for the simple match domain

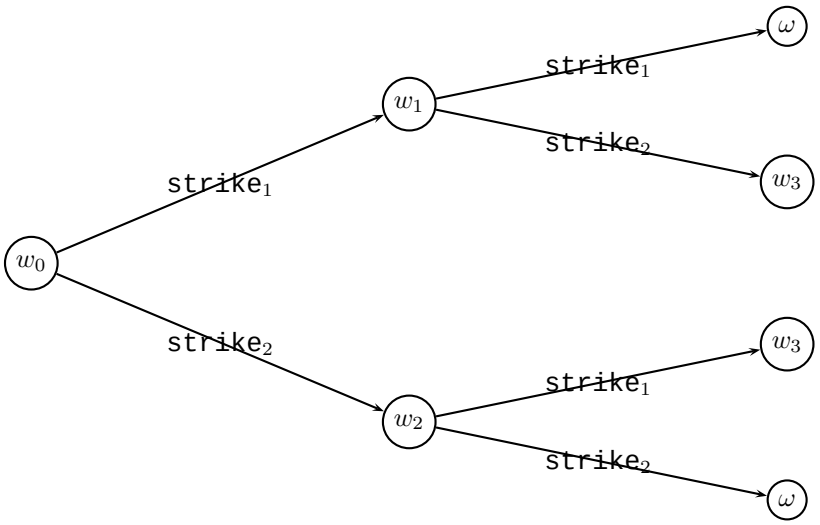

Figure 2: Selection function for the simple match domain

New fluents:	$\texttt{together}$
New Actions:	$\texttt{spill}_1, \texttt{spill}_2, \texttt{separate}$
Intended I_1:	$\{\texttt{dry}_1, \texttt{dry}_2, \texttt{oxygen}, \neg\texttt{struck}_1, \neg\texttt{struck}_2, \neg\texttt{lit}_1, \neg\texttt{lit}_2, \texttt{together}\}$
Initial condits:	$\mathit{World}[(]I_1)$
New inertial axs:	$\texttt{oxygen} \Leftarrow [L]\texttt{oxygen}$ if $\neg\texttt{oxygen} \notin L$ $\neg\texttt{oxygen} \Leftarrow [L]\neg\texttt{oxygen}$ if $\texttt{oxygen} \notin L$ $\texttt{dry}_1 \Leftarrow [L]\texttt{dry}_1$ if $\neg\texttt{dry}_1 \notin L$ and $\{\texttt{together}, \neg\texttt{dry}_2\} \notin L$
New change Axs:	$\neg\texttt{dry}_2 \Leftarrow [\texttt{spill}_2]\top$ $\neg\texttt{together} \Leftarrow [\texttt{separate}]\texttt{together}$ $\neg\texttt{dry}_1 \Leftarrow [L]\texttt{dry}_1$ if $\{\texttt{together}, \neg\texttt{dry}_2\} \subseteq L\}$.

Figure 3: Partial axioms for the extended match domain