Causal Propagation Semantics — A Study

Mikhail Prokopenko, Maurice Pagnucco, Pavlos Peppas, and Abhaya Nayak

Computational Reasoning Group, Department of Computing, Division of ICS, Macquarie University, NSW 2109, Australia mikhail.prokopenko@cmis.csiro.au, {morri, pavlos, abhaya}@ics.mq.edu.au

Abstract. A unifying semantic framework for different reasoning approaches provides an ideal tool to compare these competing alternatives. A historic example is Kripke's *possible world* semantics that provided a unifying framework for different systems of modal logic. More recently, Shoham's work on *preferential* semantics similarly provided a much needed framework to uniformly represent and compare a variety of nonmonotonic logics (including some logics of action). The present work develops a novel type of semantics for a particular causal approach to reasoning about action. The basic idea is to abandon the standard state-space of possible worlds and consider instead a larger set of possibilities — a hyper-space — tracing the effects of actions (including indirect effects) with the states in the hyper-space. Intuitively, the purpose of these hyper-states is to supply extra context to record the process of causality.
Keywords: common-sense reasoning, nonmonotonic reasoning, temporal reasoning.

In recent artificial intelligence research into reasoning about action much attention has been focussed on the role of causality [8,13]. While there is significant consensus that a causal component to reasoning systems is not explicitly necessary to solve the frame and ramification problems, it is generally considered necessary for concise solutions to these problems.

Causal theories of action have become prominent in a proliferation of reasoning about action frameworks. Each of these frameworks is couched in its own syntax and calculus for providing solutions to the frame and ramifications problems. But a cursory glance at this situation is sufficient to clearly indicate that this is grossly inadequate. What is required is an independent semantic motivation for these various proposed frameworks.

A unifying semantics would provide a basis upon which to compare the myriad approaches to reasoning about action. Moreover, it would give a clearer insight into the nature of causality underlying these various frameworks. While the prospect of a unifying semantics is a bit too ambitious for the present work we hope that some of the morals drawn may be able to serve as a first step in this direction.

One landmark proposal in the early literature on reasoning about action was Shoham's [12] *preferential semantics*. This semantics provided insight into several areas of reasoning in artificial intelligence including belief change [1] and nonmonotonic reasoning [5]. Recently Peppas *et al.* [9] have shown that it is not possible to furnish a traditional preferential style semantics for a recent causal approach to reasoning about action — McCain and Turner's causal theory of action [8]. They provided an augmented preferential semantics capable of characterising this framework. Subsequent to McCain and Turner's framework Thielscher [13] has proposed a causal approach to reasoning about action which, under certain specific conditions, subsumes McCain and Turner's approach.[1] However, this frame-

[1] There is insufficient space for us to elaborate upon McCain and Turner's approach here and to furnish a comparison with Thielscher's alternate proposal.

work is devoid of a suitable semantics. As a result, it is difficult to place this framework in perspective with competing proposals.

Put briefly, the main aim of this paper is to *furnish a semantics characterising Thieslscher's casual theory of action* . We do so by proferring a novel type of semantics. The basic idea is to abandon the standard state-space of possible worlds and consider instead a larger set of possibilities — a hyper-space — tracing the effects of actions (including indirect effects) with the states in the hyper-space. Intuitively, the purpose of these hyper-states is to supply extra context to record the process of causality.

In the following section we outline the necessary technical preliminaries for an understanding of this paper. In section 2 we briefly sketch Thielscher's causal theory of action. In Section 3 we introduce the hyper-space semantics that we shall use to characterise Thielscher's [13] approach. Section 4 will establish the necessary representation theorems. Section 5 discusses the importance of these results.

1 Technical Preliminaries

Let $\mathcal{F}$ be a finite set of symbols from a fixed language $\mathcal{B}$, called fluent names. A fluent literal is either a fluent name $f \in \mathcal{F}$ or its negation, denoted by $\neg f$. Let $L_{\mathcal{F}}$ be a set of all fluent literals defined over the set of fluent names $\mathcal{F}$. We will adopt from Thielscher [13] the following notation. If $\epsilon \in L_{\mathcal{F}}$, then $|\epsilon|$ denotes its affirmative component, that is, $|f| = |\neg f| = f$, where $f \in \mathcal{F}$. This notation can be extended to sets of fluent literals as follows: $|S| = \{|f| : f \in S\}$. By state we intend a maximal consistent set of fluent literals. We will denote the set of all states as W, and call the number m of fluent names in $\mathcal{F}$ the dimension of W. By $[\phi]$ we denote all states consistent with the sentence $\phi \in \mathcal{B}$ (i.e., $[\phi] = \{w \in W : w \vdash \phi\}$).

2 Background

The idea of minimising change in order to deduce the set of possible next states (successor states) is used quite broadly in action theories. Sometimes the notion of minimal change is defined by set inclusion (eg., PMA) [14,4,7,8], and often incorporates the frame concept or the policy of categorisation [4,7], assigning different degrees of inertia to language elements (fluents, literals, formulas, etc.). Shortcomings of particular implementations of the principle of minimal change and the policy of categorisation are well-known: imprecise or capricious definitions of minimality metrics (eg., PWA [2] vs PMA [14]), difficulties in properly categorising fluents as inertial and non-inertial, leading to increasingly complex selection mechanisms of action languages [7,10,13]), etc. These problems have generated attempts to use some notion of causality instead of or in addition to the principle of minimal change. For instance, some action theories try to embody background information in the form of domain "causal laws", pointing to the fact that, in general, propositions embracing causal dependencies are more expressive than traditional state constraints [6,8].

However, despite numerous attempts to combine a notion of causality with the principle of minimal change and/or policy of categorisation, multiple counter-examples keep reappearing, highlighting the intractability of the ramification problem. The framework suggested by Thielscher [13] criticised the categorisation policy and the principle of minimal change, arguing for the necessity of an approach based on causality. Thielscher's approach was intended to provide a method to avoid unintuitive indirect effects (ramifications), while accounting for causal relationships of a domain in hand. One of the perceived strengths of the Thielscher approach was an ability to capture not only all intuitively expected resulting states with minimal distance to the initial state, but also non-minimal solutions -

"perfectly acceptable provided all changes are reasonable from the standpoint of causality" [13]. In other words, the non-minimal solutions are those states which are reachable via causal propagation from an intermediate state. This intermediate state is determined as the nearest state to the initial state, where the direct action effects hold, while some domain constraints may be violated.

Thielscher [13] criticised minimal change on the grounds that, in his view, it rejects a potential resultant state if it is obtained by changing the values of more fluents than strictly necessary. Arguably, this skewed view of the principle is too restrictive to warrant complete abandonment of the general notion of minimal change. In this paper we question Thielscher's criticism of minimal change and contend that there is an element of minimal change at work in his framework. To demonstrate our claim we exhibit a semantics for Thielscher's causal theory of actions. This semantics can be clearly seen to employ a component of minimal change coupled with causality.

Thielscher employs two crucial notions: *action laws* and *causal relationships*. Action laws essentially describe the immediate (or direct) effects of performing an action in a given state. Causal relationships are responsible for producing the indirect effects of actions.

Thielscher employs the following notion of action specification. Each action law consists of:

- a condition C, which is a set of fluent literals, all of which must be contained in an initial state where the action is intended to be applied;
- a (direct) effect E, which is also a set of fluent literals, all of which must hold in the resulting state after having applied the action.

For simplicity, it is assumed that condition and effect are constructed from the very same set of fluent names.Therefore, the state resulting from a direct effect is obtained by simply removing set C from the initial state at hand and adding set E to it. However, execution of an action may cause further state transitions.

Definition 1. Let $\mathcal{F}$ be the set of fluent names and let $\mathcal{A}$ be a finite set of symbols called action names, such that $\mathcal{F} \cap \mathcal{A} = \emptyset$. An action law is a triple $\langle C, a, E \rangle$ where C, called condition, and E, called effect, are individually consistent sets of fluent literals, composed of the very same set of fluent names (i.e., $|C| = |E|$) and $a \in \mathcal{A}$. If w is a state then an action law $\alpha = \langle C, a, E \rangle$ is applicable in w iff $C \subseteq w$. The application of α to w yields the state $(w \setminus C) \cup E$, where $\setminus$ denotes set subtraction.

Thielscher's approach formally incorporates causal information through *causal relationships* of the form

$$\epsilon \textit{ causes } \rho \textit{ if } \Phi$$

where ϵ and ρ are fluent literals and Φ is a fluent formula based on $\mathcal{F}$, the set of fluent names.

Definition 2. Let (s, E) be a pair consisting of a state s and a set of fluent literals E. Then a causal relationship ϵ causes ρ if Φ is applicable to (s, E) iff $\Phi \wedge \neg\rho$ is true in s, and $\epsilon \in E$. Its application yields the pair (s', E'), denoted as $(s, E) \leadsto (s', E')$, where $s' = (s \setminus \{\neg\rho\}) \cup \{\rho\}$ and $E' = (E \setminus \{\neg\rho\}) \cup \{\rho\}$.

Intuitively, a causal relationship is applicable if the associated condition Φ holds, the particular indirect effect ρ is currently false, and its cause ϵ is among the current effects —

in other words, the cause has been effected, i.e., it has *changed during causal propagation* from false in the past to being true at the moment. Importantly, if the literal ϵ is not among current effects, then it is not possible to apply the causal relationship - even if ϵ is an element of a current state.

States incorporating direct action effects may violate the underlying domain constraints.[2] So, "in order to obtain a satisfactory resulting state, we compute additional, indirect effects by (nondeterministically) selecting and (serially) applying causal relationships. If this procedure eventually results in a state satisfying the domain constraints, then this is called a *successor state*" ([13]). More precisely, the set of possible successor states $Res_{\mathcal{T}}(a,\ w)$, given an initial state w and an action a, is determined as follows.

Definition 3. Let $\mathcal{F}$ be the set of fluent names, A a set of action names, $\mathcal{L}$ a set of action laws, $\mathcal{D}$ a set of domain constraints, and R a set of causal relationships. Furthermore, let w be a state satisfying $\mathcal{D}$ and let $a \in A$ be an action name. A state r is a *successor state* of w and a, $r \in Res_{\mathcal{T}}(a,\ w)$, iff there exists an applicable (with respect to w) action law $\alpha = \langle C, a, E\rangle \in \mathcal{L}$ such that

1. $((w \setminus C) \cup E, E) \overset{*}{\leadsto} (r, E')$ for some E', and
2. r satisfies $\mathcal{D}$,

where $\overset{*}{\leadsto}$ denotes the transitive closure of $\leadsto$.

As mentioned before, an occurrence of a literal ϵ in a state s does not guarantee that a causal relationship ϵ *causes* ρ *if* Φ is applicable to a pair (s, E) — to ensure applicability, the literal ϵ has to belong to the current effects E. It is interesting to note, however, that given a *transition* pair (s, E), if the literal ϵ is among current effects E, then it must be an element of the current state s. This observation can be formalised as follows.

Lemma 4. *If* $(s', E') \overset{*}{\leadsto} (s'', E'')$*, then* $E'' \subseteq s''$*.*

It is easy to observe that the set E' contains the *most recent* consistent effects that have taken place during the causal propagation $((w \setminus C) \cup E, E) \overset{*}{\leadsto} (r, E')$. In other words, although some of the effects may have been retracted from the effects set during propagation, their negations should have taken the respective places. The effects set is intended to account for both direct and indirect changes. However, it is not guaranteed that direct effects E are always preserved by the propagation. On the contrary, they can be lost (the indirect effects can be lost as well — but this obviously is less counter-intuitive).

Consider, for example, the simple action system with $\mathcal{F} = \{p, q\}$, $\mathcal{D} = \{\neg q \rightarrow \neg p\}$, $R = \{\neg q$ *causes* $\neg p$ *if* $\top\}$, and $\mathcal{L} = \{\langle\{p, q\}, a, \{p, \neg q\}\rangle\}$. The action a performed at the initial state $\{p, q\}$, results in a state $\{p, \neg q\}$. Clearly, this resultant state does not satisfy the domain constraint. The causal relationship is then applicable, whereby $(\{p, \neg q\}, \{p, \neg q\}) \leadsto (\{\neg p, \neg q\}, \{\neg p, \neg q\})$ and produces $Res_{\mathcal{T}}(a,\ \{p, q\}) = \{\neg p, \neg q\}$, where the successor state satisfies $\mathcal{D}$, while leaving one of the direct effects (p) out.

We can strengthen the concept of successor states to *conservative* successor states (denoted $Res^*_{\mathcal{T}}(a,\ S)$) as follows.

[2] The details of an algorithm translating domain constraints and the influence relation into causal relationships are described in [13].

Definition 5. Let $\mathcal{F}$, A, $\mathcal{L}$, $\mathcal{D}$, R, w, $\alpha = \langle C, a, E \rangle$ be the same as in Definition 3. A state r is a *conservative successor state* of w and a, $r \in Res^*_{\mathcal{T}}(a,\ w)$, iff

1. $r \in Res_{\mathcal{T}}(a, w)$, and
2. $E \subseteq r$.

This definition allows the causal propagation to "travel" outside the $E-$states, but mandates that it finish in a state consistent with the direct effects E.

3 Hyper-space Semantics

It has been previously argued [9] that a preferential structure with a binary relation on states demonstrates that minimal change and causality — the former captured by preferential semantics and the latter by a binary relation — together are essential in furnishing a concise solution to the frame problem. Our approach here is intended to illustrate this idea once more, now with respect to Thielscher's causal theory of action. More importantly, it is our contention that a pure preferential semantics, in the spirit of [12], cannot be obtained for causal action systems without extending the underlying language. Thus, in addition, the proposed approach may serve as another step towards a uniform preferential semantics for (extended) causal action systems.

Our intention at this stage is to consider a formalisation of action systems which faithfully captures all successor states, as defined by $Res_{\mathcal{T}}(a,\ w)$ (or $Res^*_{\mathcal{T}}(a,\ w)$), using a simpler selection mechanism. More precisely, instead of keeping an explicit (and changing) account of context-dependent action effects, we would like to use a binary (causal) relation on states. The advantage of this proposal is that a causal relation would be action-independent, unlike a history of effects. Obviously, this objective is hardly achievable without extending the action system components in some way.

Let us begin by informally describing the semantics we develop, before proceeding to establish the formal results. An expansion of the standard state-space to a hyper-space of a larger dimension generates numerous hyper-states. Any state in the standard state-space can then be associated with a number of hyper-states, creating a hyper-neighbourhood. For instance, an intermediate state (defined, for a given action and an initial state, according to Thielscher's approach) can be represented by a set of hyper-states in the expanded space. This hyper-neighbourhood will be a starting point of a propagation. An appropriately constructed binary relation on hyper-states would allow us to propagate in the hyper-space in a very simple way — without the necessity to track the causal history, and resulting in a clearly defined "final" set of hyper-states. A projection from the resulting hyper-neighbourhood back to the normal state-space would pinpoint the desired successor state of the action at hand. Intuitively, the purpose of the hyper-states is to serve as possible causal extensions of normal states, providing necessary context to the process of causal propagation. In the remainder of this section we give a formal description of this semantics.

We suggest to extend the set of fluent names $\mathcal{F}$ and incorporate more causal information in states themselves rather than rely on context-dependent causal propagation. We begin with definitions of an extended (hyper-) state. First of all, we consider a set, denoted as $\overset{\circ}{\mathcal{F}}$, of the same cardinality as the set $\mathcal{F}$, such that $L_{\mathcal{F}} \cap \overset{\circ}{\mathcal{F}} = \emptyset$. Then we define a function $j : \mathcal{F} \to \overset{\circ}{\mathcal{F}}$. Intuitively, the element $j(f)$ of the set $\overset{\circ}{\mathcal{F}}$ is an extra space-dimension, corresponding to the fluent $f \in \mathcal{F}$. Now let us consider the set $L_{\overset{\circ}{\mathcal{F}}} = \overset{\circ}{\mathcal{F}} \cup \{\neg q : q \in \overset{\circ}{\mathcal{F}}\}$. Clearly, the cardinality of the set $\overset{\circ}{L}_{\mathcal{F}}$ is equal to the cardinality of the set of fluent literals $L_{\mathcal{F}}$, and $L_{\mathcal{F}} \cap \overset{\circ}{L}_{\mathcal{F}} = \emptyset$. Another function is needed to map from $L_{\mathcal{F}}$ to $\overset{\circ}{L}_{\mathcal{F}}$, and we introduce the function $l : L_{\mathcal{F}} \to \overset{\circ}{L}_{\mathcal{F}}$, such that $l(f) = j(f)$ if $f \in \mathcal{F}$ (f is a positive literal — a fluent name), and $l(f) = \neg j(|f|)$ if $f \in L_{\mathcal{F}} \setminus \mathcal{F}$ (f is a negative literal).

The following property of the function $l(f)$ can be easily obtained.

Lemma 6. *If $f \in \mathcal{F}$, then $l(\neg f) = \neg l(f)$.*

The function $l(f)$ is intended to produce extra literals, corresponding to fluent literals in $L_{\mathcal{F}}$. We will call a literal $l(f)$ a justifier literal, and will use the abbreviation $\overset{\circ}{f}$ instead of $l(f)$ for simplicity. In addition, the set $\overset{\circ}{\mathcal{F}}$ will be referred to as the set of all justifier fluents, and the $\overset{\circ}{L}_{\mathcal{F}}$ will be referred to as the set of all justifier literals.

Having defined the function l, we can define a justifier set $\overset{\circ}{J}$ for a set of fluent literals J as $\overset{\circ}{J} = \cup_{f \in J}\{l(f)\} = \cup_{f \in J}\{\overset{\circ}{f}\}$.

Definition 7. Given a set of fluents $\mathcal{F}$, a *hyper-state* is a maximal consistent set of literals from $L_{\mathcal{F}} \cup \overset{\circ}{L}_{\mathcal{F}}$.

We will denote the set of all hyper-states as Ω, where the dimension of Ω is $2m$, m being the dimension of W. The following two functions map hyper-space Ω to normal space W and vice versa.

Definition 8. A projection from Ω to W, $p : \Omega \to W$, is the function mapping a hyper-state $s = \{f_1, ..., f_n, \overset{\circ}{f}_1, ..., \overset{\circ}{f}_n\} \in \Omega$ to a state $r = \{f_1, ..., f_n\} \in W$.

We denote the hyper-part of a hyper-state $s \in \Omega$ as $h(s) = s \setminus p(s)$. Clearly, for any $s \in \Omega$, $h(s) \cap \mathcal{F} = \emptyset$.

Definition 9. A hyper-neighbourhood of a state $r \in W$, $N : W \to 2^{\Omega}$, is the function mapping a state r to a set of hyper-states: $N(r) = \{s \in \Omega : r = p(s)\}$.

Clearly, there are 2^m states in any hyper-neighbourhood, as there are m justifier fluent names in any hyper-state allowed to vary across the neighbourhood. Intuitively, justifier literals represent explicit causes for a set $r \in W$. In other words, the set $N(r)$ is the set of states where all possible causes vary, while the (proper) literals defined on $\mathcal{F}$ are fixed. For example, given the state $r = \{a, b\}$ in the normal space W, one can consider its hyper-neighbourhood $N(r)$ containing hyper-states $\{a, b, \overset{\circ}{a}, \overset{\circ}{b}\}$, $\{a, b, \overset{\circ}{a}, \neg\overset{\circ}{b}\}$, $\{a, b, \neg\overset{\circ}{a}, \overset{\circ}{b}\}$ and $\{a, b, \neg\overset{\circ}{a}, \neg\overset{\circ}{b}\}$, where the justifier fluents $\overset{\circ}{a}$ and $\overset{\circ}{b}$ vary. Hence any subset of $N(r)$ may represent a particular causal context — the set $\{\{a, b, \overset{\circ}{a}, \overset{\circ}{b}\}, \{a, b, \overset{\circ}{a}, \neg\overset{\circ}{b}\}\}$, for instance, may correspond to a partial state $\{a, b, \overset{\circ}{a}\}$, justifying the literal $a \in r$, and leaving the literal $b \in r$ somewhat unsupported (more precisely, any *change* in a truth value of a literal will be expected to have a justification).

It is worth noting that the history component E in any causally propagated pair (s, E) cannot have more elements than m — due to the consistency of the update defined in Definition 2, as shown by Lemma 4. In a simple case, when the history component E in a pair

(s, E) has exactly m elements (or, in other words, $E = s$ by the Lemma 4), the pair can be easily represented by a single hyper-state $s \cup \mathring{s}$. For instance, the hyper-state $\{a, b, \mathring{a}, \mathring{b}\}$ can account for a causal transition pair $(\{a, b\}, \{a, b\})$. In the case when the component E has strictly less elements, $E \subset s$, the incompleteness may be represented by a partial hyper-state. A union of complete hyper-states, $\{\{a, b, \mathring{a}, \mathring{b}\}, \{a, b, \mathring{a}, \neg\mathring{b}\}\}$ can represent the pair $(\{a, b\}, \{a\})$ in a causal propagation chain where the second component carries the history of change $\{a\}$.

It is precisely the combinatorial variability of possible causes in a hyper-neighbourhood that allows us to account for different action-dependent histories in a causally propagated chain, leading to a successor state in $Res_{\mathcal{T}}(a,\ w)$. Before we formally introduce the required notion of a binary causal relation on hyper-states, let us illustrate the intention with an example.

Consider an action system with $\mathcal{F} = \{a, b, c\}$, $\mathcal{D} = \{\neg b \rightarrow \neg a\}$, $R = \{\neg b$ *causes* $\neg a$ *if* $\top\}$, and $\mathcal{L} = \{\langle \{b\}, x, \{\neg b\} \rangle\}$. Let us perform the action x at the initial state $w = \{a, b, c\}$. The action direct effect, stored in an (initial) history component, is $\{\neg b\}$, and the intermediate state is, obviously, $\{a, \neg b, c\} = (w \setminus \{b\}) \cup \{\neg b\}$. This state violates the domain constraint, but the only causal law of the system is applicable: $(\{a, \neg b, c\}, \{\neg b\}) \rightsquigarrow (\{\neg a, \neg b, c\}, \{\neg a, \neg b\})$. The state component of the yielded pair satisfies the domain constraint and therefore belongs to $Res_{\mathcal{T}}(x,\ w)$. It is easy to verify that $Res_{\mathcal{T}}(x,\ w)$ is a singleton.

Now, let us sketch how this simple propagation could be traced in the hyper-space. The hyper-neighbourhood $N(r)$ of the intermediate state $r = \{a, \neg b, c\}$ contains eight hyper-states, some of which represent the initial history component $\{\neg b\}$ — these hyper-states are precisely the states in $N(r) \cap [\neg\mathring{b}]$. The hyper-neighbourhood of the successor state $r' = \{\neg a, \neg b, c\}$ contains some hyper-states accountable for the final history component $\{\neg a, \neg b\}$. These states are precisely the states in $N(r') \cap [\neg\mathring{a} \wedge \neg\mathring{b}]$ or $N(r') \cap [\neg\mathring{a}] \cap [\neg\mathring{b}]$. Our intention, therefore, is to construct, for an action system, such a binary relation on hyper-states that a transition in hyper-space faithfully corresponds to causal propagation driven by an action-dependent history.

Formally, we define a binary relation on states in Ω as follows.

Definition 10. A binary relation $\mathcal{C}$ is defined on $\Omega \times \Omega$. We say that $C(s, s')$ if and only if there exists a causal relationship ϵ *causes* ρ *if* Φ such that

1. $p(s) \vdash \epsilon \wedge \Phi \wedge \neg\rho$
2. $h(s) \vdash \mathring{\epsilon}$
3. $p(s') = (p(s) \setminus \{\neg\rho\}) \cup \{\rho\}$
4. $h(s') = (h(s) \setminus \{\neg\mathring{\rho}\}) \cup \{\mathring{\rho}\}$

Figure 1 illustrates the generation of links by a causal relationship between hyper-states which belong to distinct hyper-neighbourhoods (the causal relationship is the same as in the example above). The fact that all the states in $N(r) \cap [\neg\mathring{b}]$ have links to the states in $N(r') \cap [\neg\mathring{a} \wedge \neg\mathring{b}]$ is not a coincidence, and will be formally captured in a definition of a successor state.

It is worth pointing out that the first condition in Definition 10 requires that the literal ϵ is a part of the $p(s)$ state — unlike Definition 2. However, Lemma 4 illustrated that this re-

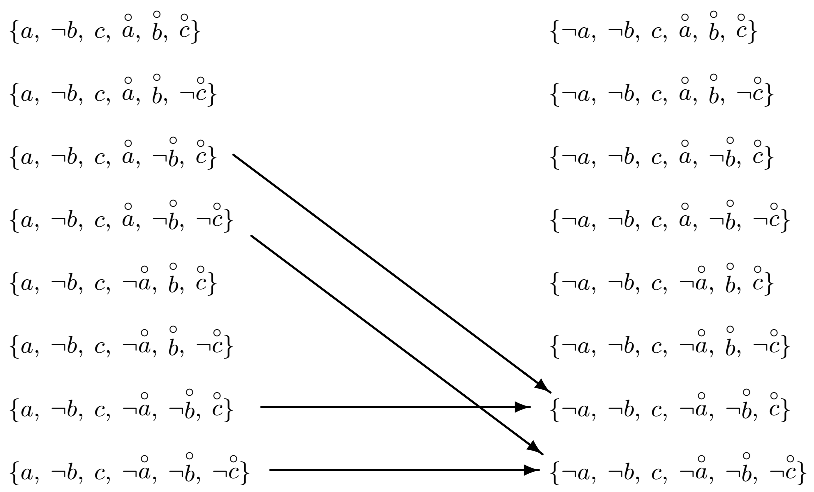

Fig. 1. The C-links between hyper-neighbourhoods of the states $\{a, \neg b, c\}$ and $\{\neg a, \neg b, c\}$, generated by a causal relationship $\neg b$ *causes* $\neg a$ *if* $\top$.

quirement is implicit in Definition 2 as well ensuring that, in this respect, the new definition is not going to be more restrictive than the former one (formally, it will be shown later).

A causal relationship ϵ *causes* ρ *if* Φ may, upon translation, generate quite a few links between hyper-states. But causal propagation expressed in terms of these links is much clearer and simpler than that of Thielscher's approach.

4 Representation Theorems

The following set will help in our analysis of causal links $C(s, s')$. Given two states $x \in W$ and $y \in W$, the set $L(x, y) = \{s \in N(x) : C(s, s'),$ for some $s' \in N(y)\}$ will be referred to as the *connection set* for the states x and y. In general, $L(x, y) \neq L(y, x)$.

An important property of the relation $\mathcal{C}$ is that there are at least 2^{m-1} links generated by one causal relationship (the minimum is attained when a causal relationship ϵ *causes* ρ *if* Φ is qualified by a complete state: $\Phi \leftrightarrow \wedge_{k=1}^{k=m} f_k$). This property leads to the following lemma.

Lemma 11. *For any two states $x \in W$ and $y \in W$, if the connection set $L(x, y) \neq \emptyset$ then there exists a justifier literal $\mathring{f}$ such that $[\mathring{f}] \cap N(x) \subseteq L(x, y)$.*

This lemma basically says that, if there is at least one $\mathcal{C}$-link between two hyper-states, then there are at least $2^{m-1} - 1$ more $\mathcal{C}$-links between hyper-states in the respective neighbourhoods, and all these links are generated by the same causal relationship.

Figure 2 illustrates the existence of a justifier literal $\neg\mathring{b}$ such that $[\neg\mathring{b}] \cap N(\{a, \neg b, c\}) \subseteq L(\{a, \neg b, c\}, \{\neg a, \neg b, c\})$.

It is possible to show that a qualified reverse observation holds as well.

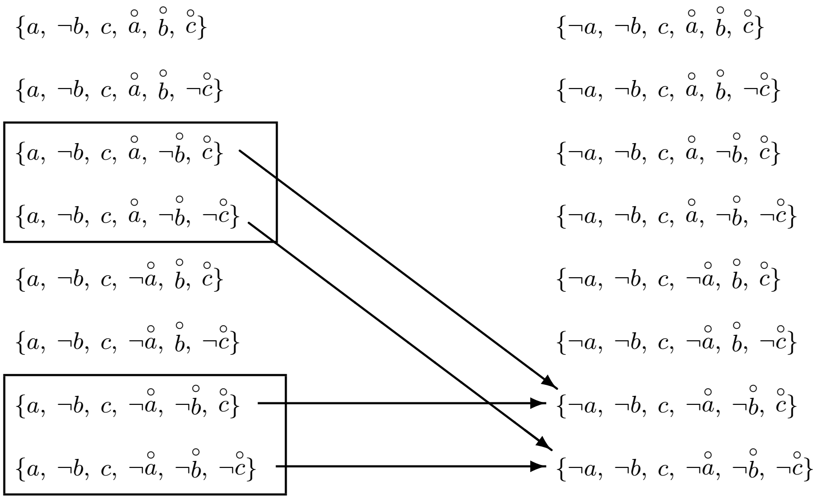

Fig. 2. All $[\neg\mathring{b}]$-states belong to the connection set$L(\{a, \neg b, c\}, \{\neg a, \neg b, c\})$.

Lemma 12. *For any two states $x \in W$ and $y \in W$, if there exists a justifier literal $\mathring{f}$ such that $[\mathring{f}] \cap N(x) \subseteq L(x, y)$, then there exists a causal relationship f causes ρ if Φ, for some Φ and where $\{\rho\} = y \setminus x$.*

The proof of this lemma progressively eliminates all literals except f, which might have been alternative causes. It capitalises on the fact that varying $m - 1$ justifier literals (having fixed $\mathring{f}$) accounts for at most $2^{m-1} - 1$ states in an hyper-neighbourhood, while there are 2^{m-1} states in the set $[\mathring{f}] \cap N(x)$.

Together, the last two lemmas show that the presence of a causal relationship underlying a $\mathcal{C}$-link is equivalent to the existence of a justifier literal $\mathring{f}$ such that $[\mathring{f}] \cap N(x) \subseteq L(x, y)$.

Corollary 13. *For any two states $x \in W$ and $y \in W$, there exists a justifier literal $\mathring{f}$ such that $[\mathring{f}] \cap N(x) \subseteq L(x, y)$, if and only if there exists a causal relationship f* causes ρ if Φ*, where $\{\rho\} = y \setminus x$, for some Φ.*

It is not surprising to observe that any connection set may not contain all $[\mathring{\epsilon}]$-states and all $[\neg\mathring{\epsilon}]$-states in any hyper-neighbourhood. Although the set R of causal relationships is allowed to include causal relationships like f *causes* ρ *if* Φ and $\neg f$ *causes* ρ *if* Φ, such ("contradictory") relationships would generate $\mathcal{C}$-links originating from different hyper-neighbourhoods. So any given hyper-neighbourhood may have outcoming $\mathcal{C}$-links generated by only one of the "contradictory" relationships. Formally, this observation is captured as follows.

Lemma 14. *For any two states $x \in W$ and $y \in W$, there is no justifier literal $\mathring{\epsilon}$ such that both $[\mathring{\epsilon}] \cap N(x) \subseteq L(x, y)$ and $[\neg\mathring{\epsilon}] \cap N(x) \subseteq L(x, y)$ hold.*

Before we define a successor state for an initial state $w \in W$ and an action a, where $\langle C, a, E\rangle$ is the action law, we need to define one more construct — a *trigger set* of hyper-states $s \in \Omega$, where the $p(s)$ state is the nearest state to w, consistent with the direct effects E, and justifier literals in $h(s)$ capture the initial (immediate) causal context.

Definition 15. A trigger set of states $\|E\|_w$ is defined for an initial state $w \in W$ and an action a, where $\langle C, a, E\rangle$ is the action law, as

$$\{s \in N(q) : q \in W, q \in min([E], \ll_w), h(s) \vdash \mathring{E}\}$$

where $x \ll_w y$ if and only if Diff$(x,\ w) \subset$ Diff$(y,\ w)$.

Here Diff$(p,\ q)$ represents the symmetric difference of p and q (i.e., $(p \setminus q) \cup (q \setminus p)$) as in PMA [14].

In other words, $\|E\|_w$ is the set contained in the hyper-neighbourhood $N(q)$ of the state q nearest to the initial state w (in terms of the PMA ordering), and the states $s \in \|E\|_w$ jointly represent the initial causally justified changes triggered by effects E. For example, consider an action law $\{\langle\{b\}, x, \{\neg b\}\rangle\}$, applied at the initial state $\{a, b, c\}$. Then the trigger set $\|\{\neg b\}\|_{\{a,b,c\}}$ contains exactly the states placed in boxes in Figure 2. The following observation can be obtained from the definition immediately.

Lemma 16. *For any initial state $w \in W$ and an action a, where $\langle C, a, E\rangle$ is the action law, $\bigcap_{s\in\|E\|_w} h(s) = \mathring{E}$.*

Intuitively, what the states $s \in \|E\|_w$ have in common in terms of justifier literals, is precisely literals in $\mathring{E}$.

Having defined a trigger set $\|E\|_w$, we can formally trace a causal propagation in the hyper-space Ω. Let $\mathcal{C}^\star$ be the transitive closure of $\mathcal{C}$.

Definition 17. We say that a hyper-neighbourhood $N(q)$, where $q \in W$, is causally triggered by the set $\|E\|_w$, denoted as $\|E\|_w \succ N(q)$ if and only if $\forall s \in \|E\|_w$, $\exists s' \in N(q)$, such that $\mathcal{C}^\star(s, s')$ holds.

A case shown previously in Figure 2 was an instance (assuming a direct action effect $\neg b$) when the trigger set $\|\{\neg b\}\|_{\{a,b,c\}}$ does trigger the hyper-neighbourhood on the right-hand side. Figure 3 gives an example when the same trigger set fails to trigger the same hyper-neighbourhood — because not all the states in the set $\|\{\neg b\}\|_{\{a,b,c\}}$ belong to the given connection set.

It is easy to check that the causal relationship which generated the connection set would not be applicable according to Thielscher's approach as well — because the cause (c) is not a part of a history component (equal to the direct effect at this stage).

Intuitively, changes triggered by the set $\|E\|_w$ propagate in hyper-space towards a hyper-neighbourhood of a possible successor state, tracing through some (causally triggered) hyper-neighbourhoods.

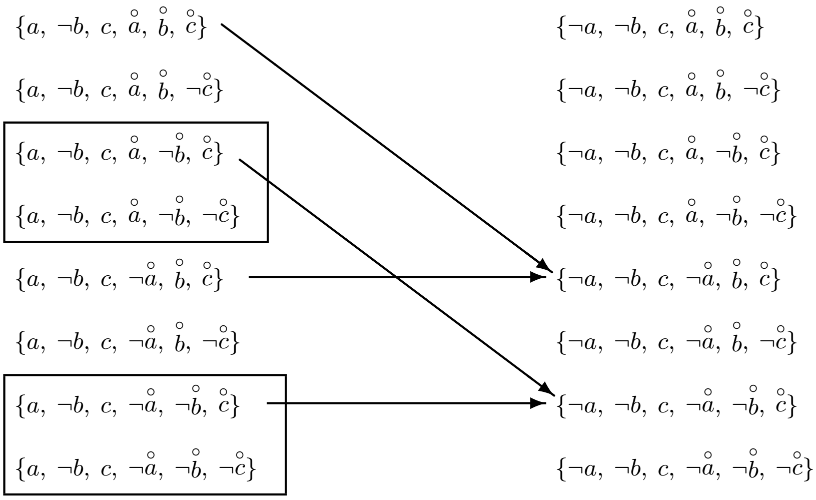

Fig. 3. The C-links between hyper-neighbourhoods of the states $\{a, \neg b, c\}$ and $\{\neg a, \neg b, c\}$, generated by a causal relationship c *causes* $\neg a$ *if* $\neg b$. Some $[\neg\mathring{b}]$-states do not belong to the connection set $L(\{a, \neg b, c\}, \{\neg a, \neg b, c\})$.

Definition 18. A state $s \in \Omega$ is *final* if and only if $\{s' : C(s, s')\} = \emptyset$. A state $r \in W$ is final if and only if $\forall s \in N(r)$, s is final.

Now we are ready to formally define a set of possible successor states $Res_\Omega(a,\ w)$ intended to faithfully capture Thielscher's resultant state set $Res_\mathcal{T}(a,\ w)$.

Definition 19. Let $\mathcal{F}$ be a set of fluent names, A a set of action names, $\mathcal{L}$ a set of action laws, $\mathcal{C}$ a causal binary relation defined by Definition 10. Furthermore, let $w \in W$ be an initial state and let $a \in A$ be an action name. A state $r \in W$ is a *successor state* of w and a, $r \in Res_\Omega(a,\ w)$, if and only if there exists an applicable (with respect to w) action law $\alpha = \langle C, a, E\rangle\ \in \mathcal{L}$ such that $\|E\|_w \succ N(r)$ and r is final.

Alternatively, $Res_\Omega(a,\ w) \equiv \{r \in W : \|E\|_w \succ N(r), r \text{ is final }\}$.

We will need the following lemma before establishing the desired representation result.

Lemma 20. *If* $\|E\|_w \subseteq N(x)$*, then* $\|E\|_w \succ N(y)$ *for some* $y \in W$ *iff* $(x, E)\ \overset{*}{\rightsquigarrow}\ (y, E')$ *for some* E'*.*

This lemma establishes a principal parallel between a propagation in the hyper-space and causal propagation of Thielscher approach.

The foregoing results now allow us to estblish the central result of this paper.

Theorem 21. $Res_\mathcal{T}(a,\ w) = Res_\Omega(a,\ w)$.

Analogous results can be obtained for conservative successor states as well if we define $Res^*_{\Omega}(a,\ w) \equiv \{r \in W : \|E\|_w \succ N(r), r \text{ is final}, E \subseteq r\}$.

Theorem 22. $Res^*_{\mathcal{T}}(a,\ w) = Res^*_{\Omega}(a,\ w)$.

5 Discussion

The semantics proposed here extends the standard state-space to a hyper-space, and works by tracing the effects of actions (including indirect effects) in the hyper-space. The hyper-states are used to supply extra context to the process of causal propagation.

These additional states are reminiscent of a semantics provided by Kraus *et al.* [5] for nonmonotonic consequence relations. It would be interesting to extract a nonmonotonic consequence relation from the result function(s) investigated here. This would facilitate a wider comparison with a wider class of logics for nonmonotonic reasoning.

While we do not prove that the strategy proposed here is capable of furnishing a semantics for approaches to reasoning about action in general, we suggest that it is a fruitful strategy to pursue in supplying a unifying semantics for a large class of such frameworks. This is, at present, the subject of ongoing investigation.

There are several questions that one is tempted to ask. Are hyper-states necessary? What is the minimum number of hyper-states required to characterise a given result function? These, too, will be left for future investigation.

References

1. Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Bradford Books, MIT Press, Cambridge Massachusetts, 1988.
2. Matthew L. Ginsberg and David E. Smith. Reasoning about action I: A possible worlds approach. Artificial Intelligence **35**:165–195, 1988.
3. Joakim Gustafsson and Patrick Doherty. Embracing occlusion in specifying the indirect effects of actions. In L. Aiello, J. Doyle, and S. Shapiro, editors, *Proceedings of the Fifth International Conference on Knowledge Representation and Reasoning*, pp. 87-98. Morgan-Kaufmann, 1996.
4. G. Neelakantan Kartha and Vladimir Lifschitz. Actions with indirect effects (Preliminary report). In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning*, pages 341–350. Bonn, 1994.
5. Sarit Kraus, Daniel Lehmann and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics, Artificial intelligence, **44**: 167–207, 1991.
6. Fangzhen Lin. Embracing causality in specifying the indirect effects of actions. In *Proceedings of International Joint Conference on Artificial Intelligence, Montreal* 1985 – 1991, 1995.
7. Vladimir Lifschitz. Two components of an action language. In *Proceedings of the Third Symposium on Logical Formalizations of Commonsense Reasoning*, Stanford, 1996.
8. Norman McCain and Hudson Turner. A causal theory of ramifications and qualifications. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1978–1984. Montreal, 1995.
9. Pavlos Peppas, Maurice Pagnucco, Mikhail Prokopenko, Abhaya Nayak and Norman Foo. Preferential semantics for causal systems. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pp. 118–123. Stockholm, 1999.
10. Mikhail Prokopenko and Pavlos Peppas. Modelling inertia in action languages (Extended Report). In G. Antoniou, A. K. Ghose and M.Truszczynski (eds.) *Inducing and Reasoning with Complex Representations*, Springer Verlag Lecture Notes in AI, Volume 1359, pages 236-249, 1998.

11. Erik Sandewall. Assessments of ramification methods that use static domain constraints. In L. Aiello, J. Doyle, and S. Shapiro, editors, *Proceedings of the Fifth International Conference on Knowledge Representation and Reasoning*. Morgan-Kaufmann, 1996.
12. Yoav Shoham. *Reasoning About Change*. MIT Press, Cambridge, Massachusetts, 1988.
13. Michael Thielscher. Ramification and causality. Artificial Intelligence **89**: 317–364, 1997.
14. Marianne Winslett. Reasoning about actions using a possible models approach. In *Proceedings of the Seventh National Artificial Intelligence Conference*, pp. 89–93, San Mateo, CA., 1988. Morgan Kaufmann Publishers.