Victor Jauregui
Maurice Pagnucco
Norman Foo

A Trajectory Approach to Causality

Abstract. In this paper we propose a new approach to address the ramification problem in common-sense reasoning about action and change. We contrast the methods of McCain and Turner [6], Thielscher [10] and Sandewall [8] and, based on some of the limitations they encounter, we introduce a *trajectory-based approach* which keeps a history of the states through which a system evolves to characterise its dynamical state. We furnish an underlying state-transition semantics and a logic that admits an expressive, dynamical account of some typical scenarios which encounter modelling difficulties in the other approaches mentioned.

Keywords: Commonsense Reasoning, Reasoning about Action and Change, Ramification Problem, Causality, Trajectory Semantics.

1. Introduction

This paper is a causal investigation into the *ramification problem* in *common-sense reasoning about action and change*. The ramification problem is concerned with accounting for *all* of the effects that result when an action is performed. In particular, when we invoke an action we reserve a certain number of effects as those for which we consider the action as being *directly* responsible. In addition, we may encounter a number of *indirect* effects which are, in some sense, removed from the principal effects which characterise the action.

Traditional approaches for addressing this problem, using background knowledge in the form of logical constraints to infer the indirect effects of an action, were observed to produce unintuitive results [2, 3, 11, 6, 10]. The challenge remains to investigate formalisms that offer accurate, concise and expressive common-sense descriptions of dynamical systems through which to capture all the intended indirect effects.

The aim of this paper is to provide a general, high-level, concise, ontologically and epistemologically expressive approach through which to model dynamical systems. To this end, in Section 2, we outline two recent approaches in the literature that will form the starting point of our analysis. In Section 3 we look at an example that exhibits a number of properties which reflect the difficulties in modelling dynamical systems.

In Section 4 we introduce the Trajectory Systems approach which generalises and extends the approaches investigated in Section 2. In Section 5 we demonstrate how the Trajectory Systems approach works by entertaining a number of examples. We proceed, in Section 6, to provide a general discus-

Studia Logica **67**: 385–401, 2001.

sion of the advantages and disadvantages of a trajectory-based perspective with respect to common-sense reasoning about action and change. We conclude, in Section 7, with a brief overview of the main ideas developed in this paper and further avenues of investigation.

2. Background

Recently a number of approaches favouring the explicit representation of *causal dependence* between parts of the domain have been proposed to overcome some of the difficulties encountered by more traditional treatments of the ramification problem (e.g., [5, 6, 10]). These approaches encode the domain dependencies of a system through descriptions which can be used to discriminate in favour of causally justified effects.

In the discussion following, on causal approaches for determining ramifications, we will use a typical propositional language in which the propositional atoms will be fluents taken from some set $\mathcal{F}$. By a *fluent* we mean a property whose truth value may change over time. We will refer to sentences in this language as *fluent formulas*. We denote a *state* by a maximal consistent set of literals from $\mathcal{F}$.

Given a set of fluent literals S, we let $\overline{S} = \{\overline{\ell} \mid \ell \in S\}$, and given a set of literals Δ, we use $S \oplus \Delta$, to denote the set $(S \setminus \overline{\Delta}) \cup \Delta$. We refer to $\oplus$ as the *override* operator[1] and say, in $S \oplus \Delta$, that Δ overrides S. For convenience, we also adopt the convention of using single literals and conjunctions of literals as sets of literals.

2.1. McCain and Turner's approach

In McCain and Turner's approach [6], causal relations are modelled as inference rules of the form:

$$\phi \Rightarrow \psi$$

where ϕ and ψ are fluent formulas. The intuition being that if we can prove ϕ then the causal dependencies of the system dictate that ψ should also hold (i.e., we can infer ψ). Note that this rule is non-contrapositive. That is, we cannot use the above rule, along with $\neg\psi$, to conclude $\neg\phi$.

Given a set of such rules C, describing the causal dependencies of the system, the remaining properties of a valid possible successor state S', follow causally from the direct effects of the action E, and those other properties which remain unaffected ($S \cap S'$). This intuition is captured by the following (from [6]):

[1] This is an adaptation of the *relational override operator* in the **Z** language.

DEFINITION 2.1. The set of all *possible successor states*, given an initial state S, a set of causal rules C, and an action with direct effect E, is the set of fixed-points S', obtained as follows: $S' \in Res_C^{MT}(E, S)$ iff

$$S' = \{\ell \mid (S \cap S') \cup E \vdash_C \ell\}$$

where ℓ is a fluent literal and $\vdash_C$ denotes derivability under propositional logic and the rules in C.

Consider the following example scenario.

EXAMPLE 2.1. (Pebble Scenario) A system, consisting of a pebble, is described by the fluents (properties):

e – the pebble is *elevated*, h – the pebble is being *held*.

If we use the set of causal rules C, to reflect the direction of causal dependence, this scenario is described by:

$$C = \{\overline{h} \Rightarrow \overline{e}\}$$

with the intuition that elevated pebbles must be held and that, in particular, unheld pebbles *become* not-elevated.

If we consider the initial state $S = \{e, h\}$, where the pebble is held aloft, and the invocation of the action of releasing the pebble, with effect $E = \{\overline{h}\}$, McCain and Turner's approach produces:

$$Res_C^{MT}(E, S) = \{\{\overline{e}, \overline{h}\}\}$$

where $\overline{e}$ is a ramification that follows causally (we infer) from C.

2.2. Thielscher's approach

Alternatively, Thielscher [10] incorporates causal information through *causal relationships* of the form:

$$\varepsilon \texttt{ causes } \varrho \texttt{ if } \Phi$$

where ε and ϱ are fluent literals and Φ is a fluent formula. The intuition above being that:

"A *change* of $\overline{\varepsilon}$ to ε *causes* a *change* of $\overline{\varrho}$ to ϱ, *provided* Φ is true."

Furthermore, to account for the change $\overline{\varepsilon}$ to ε, we maintain a set E, of those literals in which a change has been currently initiated.

Definition 2.2. Let (S, E) be a pair consisting of a state S and a set of fluent literals E. A causal relationship ε `causes` ϱ `if` Φ, is *applicable* with respect to (S, E) iff

1. $\Phi \wedge \overline{\varrho}$ is true in S, and
2. $\varepsilon \in E$.

Its application yields the pair (S', E'), where $S' = S \oplus \varrho$ and $E' = E \oplus \varrho$.

The set of possible successor states, obtained through Thielscher's [10] approach, is determined as follows.

Definition 2.3. Given a system described by the set of domain constraints B, the set of causal relationships $\mathcal{R}$, and the invocation of an action with direct effects E in state S, S' is a *successor state*, i.e., $S' \in Res_{\mathcal{R}}^{Th}(E, S)$, iff

1. $(S \oplus E, E) \overset{*}{\leadsto}_{\mathcal{R}} (S', E')$, for some E'; and
2. S' satisfies B.

Where $(S, E) \overset{*}{\leadsto}_{\mathcal{R}} (S', E')$ denotes that the pair (S', E') is obtainable from (S, E) via the application of a, possibly empty, sequence of causal relationships from $\mathcal{R}$.

Returning to our pebble of Example 2.1, we would describe it, in Thielscher's formalism, by the set of causal relationships $\mathcal{R}$, as follows:

$$\mathcal{R} = \{\overline{h} \texttt{ causes } \overline{e} \texttt{ if } \top\}$$

Given initial state S and an action with direct effects E, as before, then through the application of the rule in $\mathcal{R}$, we can again arrive at the desired result:

$$Res_{\mathcal{R}}^{Th}(E, S) = \{\{\overline{e}, \overline{h}\}\}.$$

3. Modelling dynamic systems

McCain and Turner [6] model causality as rules of inference, where causation is attributed to the presence of certain properties of the system, typically in the form of a logical expression, which indicates a condition sufficient to cause (infer) the corresponding effect in our causal relations. As the properties through which we model a system typically represent static (or quasi-static) features (e.g., alive, walking), McCain and Turner's causal rules approach makes implicit the commitment that effects follow from causes that are themselves static and localised at any instant in time.

We contrast this with Thielscher's [10] causal relationships approach which also provides a dynamic description, through the effect component

E, in the state-effect pair (S, E). Thielscher's causal relationships reflect this by requiring that the cause be a *change* from $\overline{\varepsilon}$ to ε (i.e., $\varepsilon \in E$).

Thielscher's proposal can be regarded as modelling a state-transition system in which the transitions are governed by the causal relationships characterising the system. However, the state of the underlying transition system is determined by both the static state component S, and the dynamic component E (which has no counterpart in McCain and Turner's approach), of the pair (S, E). In view of this, let us consider the following extension of Example 2.1.

EXAMPLE 3.1. (Egg Scenario) We replace the pebble of Example 2.1 with an egg and augment the scenario description with the fluent:

b – the egg is *broken*.

The intuition is that the egg, upon being released, will fall (producing $\overline{e}$), and break.

Note that here we have remodelled the original pebble scenario so as to preserve the modularity of the domain. That is, it is natural to conclude that this new domain differs from its pebble predecessor only in the fragility of the egg. We will return to the issue of modularity later.

Presently, however, our attempt to model the above system declaratively requires that we provide an additional condition which is sufficient for b. There would seem to be two candidates. The first, $e \wedge \overline{h}$, is somewhat unintuitive as we expect that our egg must have fallen (i.e., have $\overline{e}$) in order to break. The other alternative, $\overline{e} \wedge \overline{h}$, is also unintuitive, perhaps more so than the first, as it would imply that an unbroken egg placed on the ground and released would break.

Despite this, taking the first option ($e \wedge \overline{h} \rightarrow b$) we can augment the system as shown below with the corresponding causal description:

$$\begin{aligned} B &= \left\{ \begin{array}{c} e \rightarrow h \\ e \wedge \overline{h} \rightarrow b \end{array} \right\}, \\ C &= \left\{ \begin{array}{c} \overline{h} \Rightarrow \overline{e} \\ e \wedge \overline{h} \Rightarrow b \end{array} \right\}. \end{aligned}$$

If we consider the initial state $S = \{e, h, \overline{b}\}$, and the action of releasing the egg, with effect $E = \{\overline{h}\}$, we find that we are precluded from using the rule $e \wedge \overline{h} \Rightarrow b$ to infer b as, in the presence of $\overline{e}$, its antecedent will not be satisfied.

As a result we obtain the unintuitive result:

$$Res_C^{MT}(E, S) = \{\{\overline{e}, \overline{h}, \overline{b}\}\}.$$

This suggests that releasing the egg when elevated leads to it falling but not breaking, which is clearly not the result we would expect.

Following Thielscher [10], from B, we arrive at the causal description:

$$\mathcal{R} = \left\{ \begin{array}{l} \overline{h} \texttt{ causes } \overline{e} \texttt{ if } \top \\ \overline{h} \texttt{ causes } b \texttt{ if } e \end{array} \right\}$$

Because the two relationships in $\mathcal{R}$ above can be applied in different orders, this gives us:

$$Res^{Th}_{\mathcal{R}}(E, S) = \{\{\overline{e}, \overline{h}, \overline{b}\}, \{\overline{e}, \overline{h}, b\}\}$$

In the above we observe that we arrive at the desired state $\{\overline{e}, \overline{h}, b\}$. Unfortunately, we also get the other unintended state $\{\overline{e}, \overline{h}, \overline{b}\}$, obtained earlier.

Ignoring the concerns raised earlier (regarding its unintuitiveness) we find that the second alternative ($\overline{e} \wedge \overline{h} \rightarrow b$) gives us the following descriptions:

$$\begin{array}{rcl} B & = & \left\{ \begin{array}{c} e \rightarrow h \\ \overline{e} \wedge \overline{h} \rightarrow b \end{array} \right\}, \\ \mathcal{R} & = & \left\{ \begin{array}{l} \overline{h} \texttt{ causes } \overline{e} \texttt{ if } \top \\ \overline{e} \texttt{ causes } b \texttt{ if } \overline{h} \end{array} \right\} \end{array}$$

Now, despite the unintuitive underlying logical description B, we find that this causal system provides the desired result:

$$Res^{Th}_{\mathcal{R}}(E, S) = \{\{\overline{e}, \overline{h}, b\}\}.$$

We find here that despite an inaccurate static domain constraint description we have, nevertheless, correctly captured the intended dynamical behaviour of the system. In this sense the dynamic and declarative descriptions of the domain can be regarded as independent or *decoupled*.

It could be argued that it is possible to resolve the above difficulty by augmenting the description with a fluent 'f', taken to mean that the egg is *falling*, even though we might view this as unfavourable in the sense that we would like to retain as *concise* a basic model of the properties (fluents) of the system as possible. Irrespective of this, we might remodel the system with the fluent f by the following causal relationships:

$$\begin{array}{llll} \overline{h} \texttt{ causes } f \texttt{ if } e & (1) & & \\ e \texttt{ causes } f \texttt{ if } \overline{h} & (2) & \overline{e} \texttt{ causes } \overline{f} \texttt{ if } \top & (5) \\ f \texttt{ causes } \overline{e} \texttt{ if } \top & (3) & \overline{e} \texttt{ causes } \overline{f} \texttt{ if } b & (6) \\ f \texttt{ causes } b \texttt{ if } \top & (4) & & \end{array}$$

Consider first only the four relationships on the left. The problem we encounter based on these is that we have not specified the condition for the retraction of the fluent f (as we expect that on reaching the ground the egg will stop falling). One suggestion to overcome this is by making f a 'momentary' fluent with default value *false* and which, on completion of all other effects, resumes the value *false*. This, however, appears a somewhat artificial construction which itself could introduce problems with respect to how to reason about f (e.g., by reverting to being *false* it alters its value without an explicit cause; such spontaneous events are not considered in the present treatment and are not generally adaptable within the current framework).

Alternatively, we could add relationship (5), on the right, to the domain description. However, if we consider the state $S = \{\overline{e}, \overline{h}, \overline{b}, f\}$ (which we might arrive at after having released ($\overline{h}$) the egg from the initial state $S_0 = \{e, h, \overline{b}, \overline{f}\}$, and after subsequent applications of relationships (1) and (3)), we find that we have now introduced indeterminacy into the system, as the egg may break, or not, depending on whether we apply (4) or (5) first, respectively.

Another alternative might be to replace (5) with (6), ensuring that the egg must break before it stops falling, thus obtaining the desired successor (i.e., $S' = \{\overline{e}, \overline{h}, b, \overline{f}\}$). However, this is undesirable for another reason: it seems intuitively strange to establish b as necessary for $\overline{f}$. Moreover, if we consider the predecessor pebble domain, this formulation breaks the modularity that was otherwise preserved (the desire to maintain a modular description is emphasised by Denecker *et al.* [1]). Indeed, the pebble, not being fragile, is modelled in the absence of b and, hence, by the symmetries in the domains one would argue that retracting f should not depend on b. The fact that this 'correct' treatment of f requires such artificial dependencies indicates that modelling the scenario in this way may not, ultimately, be well motivated.

Another approach, which would require a slight modification of Thielscher's formalism, would be to allow multiple effects for a given causal relationship. We might then add the following into our system:

$$\overline{e}\ \texttt{causes}\ (\overline{f} \wedge b)\ \texttt{if}\ f$$

The argument for a similar use of 'complex formula' effect-rules is highlighted by Denecker *et al.* [1]. Note also that the rule has a form of negative cyclic dependency (between $\overline{f}$ and f). As we'll see later, both 'multiple effect' reformulations and cyclic dependencies characterise important properties of domains modelled by an underlying state-transition system. Thielscher's approach, however, does not allow for either multiple effects, nor for rules with direct cyclic dependencies.

The presence of cyclic rules, such as the above, suggests that the role of the cyclic fluent (f in this case) is ambiguous in the context of a system which evolves in discrete steps, such as the ones being considered here. In terms of a declarative state constraint description, the fluent f must always be false as it must start and end with the value *false*. In this respect, f identifies a transient property whereby, with respect to such a fluent, the system enters a kind of turbulent dynamical state. This 'turbulence' seems to provide the source of the difficulty in modelling such systems on static constraints when using state-transition based approaches.

The philosophy behind Thielscher's approach is that causal chains propagate so as to return the system to a state that preserves the integrity of the domain constraints (see Definition 2.3). This can be interpreted as the system entering an *unstable* state through the application of an action, whereupon the causal forces dictate how the system resolves to a *stable* final state (i.e., a successor that satisfies the domain constraints).

We argue that the fundamental problem encountered in the above example is that of attempting to describe a dynamical property of the system (i.e., stability with respect to falling) within a static context. That is, given the specification above, a declarative description of the system is not sufficient to adequately determine the stability of a state.

For example, the state $S = \{\overline{e}, \overline{h}, \overline{b}\}$ might be regarded as stable if we had arrived at it by releasing the egg while on the floor (i.e., following a transition of h to $\overline{h}$). On the other hand, if we had arrived at the state from a state in which the egg was elevated (i.e., following a transition of e to $\overline{e}$), we might concur that the state is the unstable result of the egg falling, thereby causing our system to proceed to a state in which the egg is also broken.

The idea that the rules that govern indirect effects need not, in general, be coupled with a system's domain constraint description is also entertained by Denecker *et al.* [1]. Note that this coupling, which we will also waive, is strongly adhered to by Thielscher, both in the sense that his causal relationships are motivated by their intended use to 'correct' unstable states, as well as to determine possible final successor states (see Definition 2.3).

In summary, we can conclude that, from the considerations expressed above, with respect to the example provided, there is evidence in support of the potential benefits to be gained from a different approach to dynamical modelling.

4. Trajectory systems

Based on the motivations above, our task in this section is to develop, generalise and formalise the dynamic characteristics of a typical state-transition system based approach, in the spirit of Thielscher [10] and Sandewall [8], for determining the indirect effects of actions.

As we saw earlier, Thielscher incorporates dynamics into his system through the set of changed fluents E, in the state-effect pair (S, E), and semantical considerations regarding the application of causal relationships. Presently, we argue that such dynamics can be interpreted intuitively, and more fundamentally, in the context of the sequence of states through which a state-transition system evolves. From this perspective, for example, a change in a fluent can be identified with the pair of successive states in the sequence for which the fluent assumes different values (e.g., ε in one and $\overline{\varepsilon}$ in the other).

Note that we have remained vague as to the interpretation of a trajectory with respect to whether state transitions are intended to evolve in time or occur instantaneously and/or simultaneously. At this point we need not make a distinction either way, thus allowing for both interpretations. We will return to this issue later.

Moreover, as in Denecker *et al.* [1], we replace, at a high level, a static state constraint description of the system with a dynamical one in which the evolution of the system in response to a given action is maintained. Fundamental to such an analysis is the following definition.

DEFINITION 4.1. A *trajectory* is a sequence of states, which we denote by the expression $t_n = \langle S_0, \ldots, S_n \rangle$, for $n \geq 0$.

Henceforth, trajectories will form the operational basis in our discussion of dynamical systems. Note that the use of trajectories has been considered before by, for example, Sandewall [7, 8].

4.1. Expressing knowledge about trajectories

Syntax. We now propose a simple high-level language sufficiently expressive to enable us to make statements about the truth of parts of the current state of the system, and the states that preceded it.

DEFINITION 4.2. We obtain a *trajectory formula* by taking a typical phrase structure grammar for propositional logic and adding the production:

$$X \to X \rightsquigarrow X$$

where X stands for a trajectory formula.

Semantics. The intuition that we want to capture is that the expression $\phi \rightsquigarrow \psi$ holds with respect to a certain trajectory iff ψ currently holds and ϕ was true in the preceding state of the system in a trace (trajectory) of how the system evolved up to the present state. We define this in the following.

Definition 4.3. Let $t_n = \langle S_0, \ldots, S_n \rangle$ be a trajectory, f be a propositional fluent and ϕ and ψ be trajectory formulas. We define that a formula ϕ holds with respect to trajectory t_n, denoted by $t_n \models \phi$, as follows:

$$
\begin{array}{lll}
t_n \models f & \text{iff} & S_n \models f \\
t_n \models \neg\phi & \text{iff} & t_n \not\models \phi \\
t_n \models \phi \wedge \psi & \text{iff} & t_n \models \phi,\ t_n \models \psi \\
 & \vdots & \\
t_n \models \phi \rightsquigarrow \psi & \text{iff} & t_n \models \psi,\ t_{n-1} \models \phi \qquad n > 0
\end{array}
$$

with the convention that the connective $(\rightsquigarrow)$ is left associative.

Note, for a literal ℓ and a state S, $S \models \ell$ iff $\ell \in S$. Also note that by nesting $(\rightsquigarrow)$, e.g., $\phi \rightsquigarrow \psi \rightsquigarrow \chi$ (which means $(\phi \rightsquigarrow \psi) \rightsquigarrow \chi$), we can refer to similar conditions and effects prevailing more than one step in the past.

Notational extensions. Though $(\rightsquigarrow)$ is the main primitive for describing dynamics in our trajectory systems semantics, we will often find more useful the following variants.

$$
\left.
\begin{array}{lll}
t_n \models \phi \overset{k}{\rightsquigarrow} \psi & \text{iff} & t_n \models \psi,\ t_{n-k} \models \phi \\
t_n \models \phi \overset{*}{\rightsquigarrow} \psi & \text{iff} & (\exists k)\ t_n \models \phi \overset{k}{\rightsquigarrow} \psi
\end{array}
\right\} \ 0 \le k \le n
$$

4.2. State-transition semantics

Following both Thielscher [10] and Sandewall [8], we develop a state-transition semantics through which to compute indirect effects. Our task is to incorporate knowledge, expressed using the above logic (which, following Sandewall [8], we will refer to as our *base logic*), through which to capture the intended dynamics of our system.

Definition 4.4. We represent a *transition rule* as an expression of the form:

$$\phi \Rightarrow \psi$$

where ϕ is a *trajectory formula* and ψ is a fluent formula. We refer to a set of transition rules $\mathcal{C}$, as a *causal system.*

Definition 4.5. A transition rule $\phi \Rightarrow \psi$ is *applicable* with respect to the trajectory t_n iff $t_n \models \phi$ and $t_n \models \neg\psi$.

Definition 4.6. Given a state S and an action with direct effects E, the *invocation* of the action in the initial state S yields the trajectory:

$$t = \langle S, S \oplus E \rangle$$

We noted earlier that it is the trajectory that captures a system's dynamics. This is reflected by the following.

DEFINITION 4.7. We denote that trajectory $t_n = \langle S_0, \ldots, S_n \rangle$ *yields* trajectory t' *through* transition rule $\phi \Rightarrow \psi$ by:

$$t_n \rhd_{\phi \Rightarrow \psi} t'$$

We define this operation through:

$$t' = \begin{cases} t_n \circ \langle S_n \oplus \psi \rangle & \text{if } \phi \Rightarrow \psi \text{ is applicable wrt. } t_n \\ t_n & \text{otherwise} \end{cases}$$

where '$\circ$' denotes concatenation. We define the corresponding operation $\overset{*}{\rhd}$, for a sequence (or chain) of rules $\sigma = \langle r_1, \ldots, r_k \rangle$, as follows:

$$\begin{array}{lll} t \overset{*}{\rhd}_{\epsilon} t & & \text{where } \epsilon = \langle\rangle \\ t \overset{*}{\rhd}_{\sigma \circ \langle r \rangle} t'' & \text{iff} & t \overset{*}{\rhd}_{\sigma} t' \rhd_r t'' \quad \text{for some } t' \end{array}$$

where $t \overset{*}{\rhd}_{\sigma} t' \rhd_r t''$ means $t \overset{*}{\rhd}_{\sigma} t'$ and $t' \rhd_r t''$.

That is, only transition rules that are applicable lead to a state change, modifying the current trajectory by appending the ensuing state onto the end of the old trajectory.

We determine the possible successor states of an action as follows.

DEFINITION 4.8. We define the set of *possible successors* (wrt. $\mathcal{C}$) of an action with effects E, invoked in state S, by $Res^*_{\mathcal{C}}(E, S)$; where we have:

$$T_{\mathcal{C}}(E, S) = \{t' \mid (\exists \sigma)\ t \overset{*}{\rhd}_{\sigma} t',\ (\forall r \in \mathcal{C})\ t' \rhd_r t'\}$$

where $t = \langle S, S \oplus E \rangle$.

$$Res^*_{\mathcal{C}}(E, S) = \{S' \mid (t \circ \langle S' \rangle) \in T_{\mathcal{C}}(E, S)\}.$$

In Definition 4.8, $T_{\mathcal{C}}(E, S)$ is the set of all *stable* trajectories t' — stability being indicated by $(\forall r \in \mathcal{C})\ t' \rhd_r t'$ — that are *reachable* from the initial state S, i.e., $(\exists \sigma)\ t \overset{*}{\rhd}_{\sigma} t'$, through the application of the action with direct effects E and the rules in $\mathcal{C}$. From this we obtain the final state of each such trajectory as the desired successor state.

5. Application of trajectory systems

Having described the Trajectory Systems approach, we now demonstrate the expressiveness of the method through the treatment of a number of examples.

EXAMPLE 5.1. (Egg Scenario – revisited) Returning to this example, we claim that it is the *event* of *falling*, which we can model by $(e \rightsquigarrow \overline{e})$, in conjunction with $\overline{h}$, that is responsible for the egg breaking. That is, we might describe this scenario, in our Trajectory Systems approach, by the causal system:

$$\mathcal{C} = \left\{ \begin{array}{c} (e \rightsquigarrow \overline{e}) \wedge \overline{h} \Rightarrow b \\ \overline{h} \Rightarrow \overline{e} \end{array} \right\}$$

If we consider the initial state $S = \{e, h, \overline{b}\}$, where the unbroken egg is held aloft, and the invocation of the action of releasing the egg, with effect $E = \{\overline{h}\}$, we obtain the trajectory:

$$t_1 = \langle \{e, h, \overline{b}\}, \{e, \overline{h}, \overline{b}\} \rangle$$

Subsequent application of the second and first rules respectively yields:

$$t_2 = t_1 \circ \langle \{\overline{e}, \overline{h}, \overline{b}\} \rangle, \qquad t_3 = t_2 \circ \langle \{\overline{e}, \overline{h}, b\} \rangle$$

Since no more rules are applicable t_3 is stable with respect to the system $\mathcal{C}$. Hence, $\{\overline{e}, \overline{h}, b\} \in Res^*_{\mathcal{C}}(E, S)$, is the only possible successor. Note that we did not have to augment the domain description (e.g., by adding a fluent f) to achieve the desired result. In this respect, we avoid the need to augment the domain description (say, with a fluent f) and its inherent difficulties, and yet provide a faithful dynamical representation of the system.

Moreover, if we consider the initial state $S = \{\overline{e}, \overline{h}, \overline{b}\}$, where the egg rests on the floor unheld, and we consider the action of throwing the egg up in the air, with effect $E = \{e\}$, we would expect the egg to become elevated (e), come down ($\overline{e}$) and break (b). This intuition would be captured by invoking the following trajectory:

$$t'_1 = \langle \{\overline{e}, \overline{h}, \overline{b}\}, \{e, \overline{h}, \overline{b}\} \rangle$$

If we look at the above system $\mathcal{C}$, we observe that, as we might expect, it responds identically to t_1 and t'_1 despite the fact that $t_1 \neq t'_1$, thus allowing us to obtain the same result $Res^*_{\mathcal{C}}(E, S) = \{\{\overline{e}, \overline{h}, b\}\}$, in both cases.

This situation, however, cannot be captured by either of the approaches of McCain and Turner [6] or Thielscher [10]. McCain and Turner's approach does not allow for the retraction of its direct effects. Such systems, where an effect is subsequently retracted, are called *non-respectful* by Sandewall [8], and characterise domains in which indirect effects are intended to proceed simultaneously. By accommodating both simultaneous and time evolving effects within the same formalism we need impose no such restriction. Likewise, Thielscher demands no such limitation, however, in his approach we cannot (without modification of the formalism) obtain the causal relationship e `causes` $\overline{e}$ `if` $\top$, which would be required (as, in his formalism, causes must be changes) in this instance to model the system correctly.[2]

EXAMPLE 5.2. (Parallel Switches Scenario)

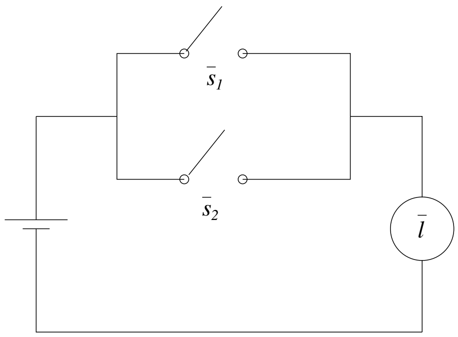

Figure 1. Parallel Switches Scenario.

Consider the domain depicted in Figure 1, described by both Sandewall [8] and Thielscher [10]. We would describe this scenario by the fluents:

s_1 – *switch 1* is closed, s_2 – *switch 2* is closed, l – the *light* is on.

This example would have the domain description:

$$B = \{s_1 \vee s_2 \leftrightarrow l\}$$

That is, the two switches s_1 and s_2 are connected in parallel and the light will turn on if and only if either of the switches is closed. Consider the initial state $S = \{\overline{s}_1, \overline{s}_2, \overline{l}\}$, in which neither switch is closed, and the action of turning the light on, with effect $E = \{l\}$. This action is intrinsically non-deterministic as either or both of the switches could close to satisfy the effects in E.

We note, however, that through Thielscher's proposal we can only obtain the relationships:

$$\mathcal{R} = \left\{ \begin{array}{l} l \texttt{ causes } s_1 \texttt{ if } \overline{s}_2 \\ l \texttt{ causes } s_2 \texttt{ if } \overline{s}_1 \end{array} \right\}$$

Unfortunately, these do not allow us to capture the alternative in which both switches close simultaneously. This limitation is due to the fact that Thielscher's causal relationships effect only a single fluent change, which,

[2] Based on his method of extracting causal relationships from the constraints B and a relation of potential influence between pairs of fluents $\mathcal{I}$.

as we saw earlier when we discussed the egg scenario for the first time in Section 3, is a restriction we would, in general, wish to waive (i.e., to allow a transition to produce multiple effects based on complex formulas).

By relaxing the single change restriction we can, in the Trajectory Systems framework, describe this scenario by:

$$\mathcal{C} = \left\{ \begin{array}{lcl} (\bar{l} \overset{*}{\rightsquigarrow} l) \wedge \overline{s}_1 \wedge \overline{s}_2 & \Rightarrow & s_1 \\ (\bar{l} \overset{*}{\rightsquigarrow} l) \wedge \overline{s}_1 \wedge \overline{s}_2 & \Rightarrow & s_2 \\ (\bar{l} \overset{*}{\rightsquigarrow} l) \wedge \overline{s}_1 \wedge \overline{s}_2 & \Rightarrow & s_1 \wedge s_2 \end{array} \right\}$$

If we look at the form of the above transition rules we note that they share the same antecedent (i.e., they are of the cyclic form discussed earlier in which one fluent is responsible for negating itself). In the context of trajectories, however, due to their dynamical character, we do not suffer the same manifestation of inherent modelling problems. Instead we find that, if one is applicable the others will be also. Yet, the application of any of the rules prevents the others from being applicable subsequently, permitting us to obtain any of the desired successors. That is:

$$Res^*_{\mathcal{C}}(E, S) = \left\{ \begin{array}{c} \{s_1, \overline{s}_2, l\} \\ \{\overline{s}_1, s_2, l\} \\ \{s_1, s_2, l\} \end{array} \right\}$$

6. Discussion

In the previous sections we motivated, defined and applied, to some typical examples, the method of trajectories for common-sense reasoning about action and change. In particular, the trajectory approach described is useful for the following reasons.

Generality. The Trajectory Systems approach allows us to generalise the notion of a dynamical system, modelled as a state-transition system in the spirit of Thielscher [10] and Sandewall [8], by taking the trajectory as the fundamental descriptive and operational entity for deriving the dynamical properties of a domain. In this sense we can obtain the following important result with respect to the Trajectory Systems approach and Thielscher's formalism.

THEOREM 6.1. *For a system of causal relationships $\mathcal{R}$, we can obtain a corresponding system of transition rules $\mathcal{C}$, for which:*

$$Res^{Th}_{\mathcal{R}}(E, S) \subseteq Res^*_{\mathcal{C}}(E, S).$$

That is, for any system of causal relationships $\mathcal{R}$, we can obtain an equivalent system of transition rules $\mathcal{C}$, such that any possible successor obtained through Thielscher's approach can be captured through the Trajectory Systems approach.

Furthermore, the trajectory language described can be regarded as an abstract dynamical description of a system. In this sense it should be amenable to comparison with other formalisms used for addressing the ramification problem (e.g., the formalisms of Shanahan [9], Denecker *et al.* [1], Gustafsson & Doherty [4], Lin [5]).

In general terms we then claim that: *The main advantage of the trajectory approach is that we can deduce the dynamic aspects of the domain description by maintaining the history of static states through which our system evolves and extrapolating a dynamical analysis based on this.* This is beneficial for a number of reasons.

Firstly, it allows us to simplify the description of the dynamic system state. We need not, as Thielscher [10] does with the set of effects E, in the pair (S, E), maintain a separate component to explicitly model the dynamical aspects of the system. We also eliminate the need to introduce additional fluents (such as the fluent f in the egg example) to resolve dynamical ambiguities that arise due to our attempts to adhere to static domain constraints. Furthermore, we follow Denecker *et al.* [1] in arguing for the need to decouple the causal rules from the declarative domain description.

Secondly, by using an underlying state-transition semantics, in the spirit of Thielscher [10] and Sandewall [8], we inherit the powerful features that such systems exhibit, which, as was shown in the Parallel Switches Scenario of Example 5.2, allow us to model non-deterministic, simultaneous and concurrent effects in a uniform setting.

Expressiveness. As we suggested in the preceding discussion, the methods of McCain and Turner [6] and Thielscher [10] make implicit ontological commitments regarding the nature of the causal relation. Based on these we can reduce most causes to:

- *Facts* (as in [6] & [8]); and/or
- *Events* (as in [10])

We argue that for a significant class of scenarios, causes typically take the form of either of the above. We can model events as assertions about the system that are distributed across a trajectory (e.g., the event of the egg *falling* was modelled by the trajectory formula $e \leadsto \overline{e}$) without the need to introduce additional fluents to our basic domain description. This makes the Trajectory Systems approach expressive enough to account for, and model, both the above types as causes, providing us with a broad ontology through which to describe dynamical systems.

Also, as we have seen, the proposed approach is *versatile* in the sense that it allows us to model indirect effects that are either sequential in a time line, simultaneous (instantaneous) or concurrent. Though such ambiguity may arguably be problematic in certain contexts it offers the advantage of establishing a common framework through which to model systems exhibiting a broad range of dynamical properties.

7. Conclusion

We have attempted, in this paper, to develop the ideas of McCain and Turner [6], Thielscher [10] and Sandewall [8], and augment them, given some of the challenges encountered, to produce the Trajectory Systems approach for addressing the ramification problem in common-sense reasoning about action and change.

We argued for the *generality* and *versatility* of the trajectory approach by demonstrating, primarily through the Egg Scenario example, the limited expressiveness encountered by the other approaches mentioned above and how the Trajectory Systems approach, using a trajectory for dynamic context, overcame such difficulties as cyclic dependencies, systems not amenable to the coupling of state constraints with causal effects, and concurrent and simultaneous effects.

Moreover, we demonstrated that the Trajectory Systems approach offers a *uniform, concise, ontologically and epistemologically expressive* approach that allows us to model systems dynamically, through the intermediate states of causal chains, where we otherwise encounter difficulties in providing an adequate faithful static description.

In concluding, this work suggests commonalities with other approaches towards describing dynamic environments and ramifications. In this respect, the next logical step is to investigate the extent to which the ideas in this paper can be reconciled, translated or reformulated with respect to such formalisms in the literature, e.g., Shanahan [9], Denecker *et al.* [1], Gustafsson & Doherty [4], Lin [5].

Acknowledgements

The authors would like to thank Mary-Anne Williams, chair of the 3rd Australian Commonsense Reasoning Workshop organised in conjunction 12th Australian Joint Conference on Artificial Intelligence (AI'99), hosted at the University of New South Wales. Furthermore, we would like to thank John McCarthy, Michael Thielscher and all the other participants for their contributions.

References

[1] Denecker, M., D. Dupré, K. Belleghem, 'An inductive definition approach to ramifications', in *Linköping Electronic Articles in Computer and Information Science*, Linköping University Electronic Press, 3, 1998.

[2] Ginsberg, M., D. Smith, 'Reasoning about action I: A possible worlds approach', *Artificial Intelligence* 35, 1988.

[3] Ginsberg, M., D. Smith, 'Reasoning about action II: The qualification problem', *Artificial Intelligence* 35, 1988.

[4] Gustafsson, J., P. Doherty, 'Embracing occlusion in specifying the indirect effects of actions', in *Proceedings of the Fifth International Conference on Knowledge Representation and Reasoning*, Morgan-Kaufmann, 1996.

[5] Lin, F., 'Embracing causality in specifying the indirect effects of actions', in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pp. 1985-1991, 1995.

[6] McCain, N., and M. Turner, 'A causal theory of ramifications and qualifications', in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pp. 1978-1984, Montreal, 1995.

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

[8] Sandewall, E., 'Transition cascade semantics and first assessments results for ramification', Preliminary report, IDA Technical Report, 1996.

[9] Shanahan, M., 'The ramification problem in the event calculus', in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 1999.

[10] Thielscher, M., 'Ramification and causality', *Artificial Intelligence* 89:317-364, 1997.

[11] Winslett, M., 'Reasoning about actions using the possible models approach', in *Proceedings of the Seventh National Artificial Intelligence Conference*, San Mateo, CA., 1988

Victor Jauregui and Norman Foo
Knowledge Systems Group
School of Computer Science and Engineering
University of New South Wales
NSW 2052, Australia
{vicj,norman}@cse.unsw.edu.au

Maurice Pagnucco
Computational Reasoning Group
Department of Computing
Division of ICS
Macquarie University
NSW 2109, Australia
morri@ics.mq.edu.au