Primary and Secondary Events

John Bell
Applied Logic Group
Department of Computer Science
Queen Mary and Westfield College
University of London
London E1 4NS
jb@dcs.qmw.ac.uk

August 18, 2000

Abstract

A formal, logical, theory of events is developed and used as the basis for a definition of causation and to provide a pragmatics for causal counterfactuals.

The theory begins with with a logical formalization of events as represented in the planner STRIPS. The resulting *inertial theories* include a common sense law of inertia and their pragmatics is based on the principle of chronological minimization. The theory of events is then developed by removing some of the simplifying assumptions of the STRIPS representation.

To begin with the assumption that events always succeed, given that their preconditions are true when they occur, is dropped. Instead a common sense law of change is introduced with the effect that events normally succeed but may fail if some qualification (some unusual or unexpected condition) arises. The pragmatics of the resulting *causal theories* is provided by refining the pragmatics of inertial theories. In particular, where there is a conflict between change and inertia, preference is given to change.

The resulting defeasibility of events makes it possible to remove the STRIPS assumption that only one event occurs at any given time; as conflicting events now result in failure rather than inconsistency. However it is desirable to provide a means for adjudicating between conflicting events. Consequently the theory is extended so that qualification preferences can be specified and the pragmatics are refined accordingly.

A further STRIPS assumption is that the effects of an event are invariant. In order to remove this assumption a distinction is drawn between primary events, events of the kind described above, and secondary events. Secondary events are invoked by primary events in appropriate contexts, they are causally dependent on primary events and when successful produce the additional, variant, effects of the primary events which invoke them. As a result, both the context-dependent effects of events (their ramifications) and non-deterministic events (events whose outcome is uncertain) can be represented.

The theory of primary and secondary events is then used as the basis for a definition of causation which can be seen as providing a formal account of Mackie's definition of a cause as an INUS condition.

A major theme of this paper is the conflict between the opposite but complementary principles of change and inertia. For much of the paper the emphasis is on restricting the effects of inertia and favouring change. The resulting imbalance becomes unnatural when considering the combined effects of several events. However the definition of causation makes it possible to state a law of change and inertia, which is used to regulate change and inertia and thus ensure a proper balance between them: while *caused* change is preferred to inertia, inertia is preferred to *uncaused* change.

The theory is then used to provide a pragmatics for causal counterfactuals, thereby extending the representation of actual causal reasoning to counterfactual causal reasoning. This is illustrated by the use of causal counterfactuals in order to obtain explanations.

1 Introduction

In this paper a formal, logical, theory of events is developed and used as the basis for a definition of causation and to provide a pragmatics for causal counterfactuals.

The theory is expressed in the Temporal Calculus, $\mathcal{TC}$, which is defined in Section 2. The section begins with an informal overview which is intended to provide a sufficient introduction for readers wishing to get quickly to the heart of the paper. The formal syntax and semantics of $\mathcal{TC}$ are then presented.

The theory of events begins in Section 3 with with a logical formalization of events as represented in the planner STRIPS. The resulting *inertial theories* include a common sense law of inertia and their pragmatics is based on the principle of chronological minimization. As the discussion it is aimed at the general reader, specialist readers will find much this material familiar. However, proceeding slowly and carefully also has the advantage that it is possible to be clear about the origins of the present work and about the basic intuitions and assumptions on which subsequent developments are based.

The theory of primary events is then developed by removing the assumption that that events always succeed given their preconditions. Instead a common sense law of change is introduced with the effect that events normally succeeed, but may fail if some qualification (some unusual or unexpected condition) arises. The pragmatics of the resulting *causal theories* is provided by refining the pragmatics of inertial theories. In particular, where there is a conflict between change and inertia, preference is given to change.

The resulting defeasibility of events makes it possible to remove the STRIPS assumption that only one event occurs at any given time; as conflicting events now result in failure rather than inconsistency. However it is desirable to provide a means for adjudicating between conflicting events. Consequently the theory is extended so that qualification preferences can be specified and the pragmatics are refined accordingly.

A further STRIPS assumption is that the effects of an event are invariant. In order to remove this assumption a distinction is drawn in Section 4 between primary events, events of the kind described above, and secondary events. Secondary events are invoked by primary events in appropriate contexts, they are causally dependent on primary events and when successful produce the additional, variant, effects of the primary events which invoke them. As a result, both the context-dependent effects of events (their ramifications) and non-deterministic events (events whose outcome is uncertain) can be represented.

A major theme in the discussion is the restriction of the use of the axiom of inertia. Where its use conflicts with the axiom of change, preference is given to the axiom of change. Moreover, as the examples given in Section 4 show, using the inertia axiom together with domain axioms to reason about change leads to unintuitive consequences. The addition of appropriate secondary events in these examples can thus be seen as providing the additional causal element which overcomes the ill effects of the inertia axiom.

More generally, it is methodologically undesirable to use the inertia axiom and domain axioms to reason about change. Consequently Section 5 proposes a threefold strategy for representing reasoning about change and inertia, according to which: all change is represented by means of causal axioms; the domain axioms are used to define the domain and to maintain its integrity; and the inertia axiom is confined to its proper task, the representation of inertia. However in order to realize this strategy fully it is necessary to moderate the ascendancy of the principle of change, and in order to do so an account of causation is needed.

This is provided in Section 6 where the theory of primary and secondary events is used as the basis for a definition of causation which can be seen as a formalization of Mackie's suggestion that a cause is an INUS condition; that a cause is an insufficient but necessary part of an unnecessary but sufficient condition for its effect. The definition of causation is reductive, in that it employs the more primitive notions of change and inertia discussed earlier, and consequently the definition is of interest in its own right. Moreover it is put to work immediately in the statement of a Law of Change and Inertia, which can be used to regulate change and inertia and thus ensure a proper balance between them: while *caused* change is preferred to inertia, inertia is preferred to *uncaused* change.

The resulting theory can be used to provide the pragmatics for causal counterfactuals. An indication of how this is done is given in Section 7, where it is shown how these counterfactuals can be used in order to obtain explanations.

Related work is discussed in Section 8, and extensions to the proposed theory are discussed briefly in the concluding section.

2 The Temporal Calculus

The formal language which is used in this paper is called the Temporal Calculus, or simply $\mathcal{TC}$. $\mathcal{TC}$ is so called because it combines features of temporal logic with features analogous to those of the Situation Calculus [28, 29]. In this section an informal introduction to $\mathcal{TC}$ is given and this is followed by the formal syntax and semantics.

2.1 Informal introduction to $\mathcal{TC}$

$\mathcal{TC}$ is based on Kleene's strong three-valued language [19]. Kleene introduced the third truth value *undefined* in order to be able to describe computations which may not, in principle, terminate. The language can also be given an epistemic, resource-bounded, interpretation. Typically we have incomplete, or partial, information and our reasoning resources are limited. Consequently we cannot, as the classical ideal suggests, determine the truth or falsity of every sentence. Rather, we concentrate our efforts on attempting to establish the truth or falsity of those sentences that we consider relevant and ignore, treat as undefined, the truth values of the rest. Accordingly, a sentence can be true (established by the reasoner(s) as being true), false (established by the reasoner(s) as being false) or undefined (ignored by the reasoner(s) as irrelevant, or unestablishable within the reasoner(s) resource limitations). Appropriately, the truth conditions of the logical constants yield a Boolean truth value wherever possible. For example, the sentence $\neg\phi$ is true if ϕ is false, false if ϕ is true, and undefined otherwise. Similarly, the sentence $\phi \wedge \psi$ is true if ϕ and ψ are both true, false if either is false, and undefined otherwise. The analogues of classical disjunction, $\vee$, and material implication, $\supset$, can be defined in the usual way. Furthermore, the language is readily extended to the first-order case. Atomic sentences are interpreted using partial functions. Thus an atomic sentence of the form $r(u_1, \ldots, u_n)$, a *Kleene atom*, may be true, false or undefined in a model. A universally quantified sentence $\forall v \phi$ is then true if every instance of ϕ is true, false if one instance is false, and is undefined otherwise. The existential quantifier, $\exists$, can also be defined in the usual way.

In order to increase the expressivity of Kleene's language the *undefined* operator, ?, is added. Informally the sentence $?\phi$ states that the truth value of ϕ is undefined; that is, that neither the truth nor the falsity of ϕ is established. The sentence $?\phi$ is thus classically valued. The undefined operator can be used to define *weak affirmation*, *weak negation*, and *defined* operators, making it possible to express meta-level distinctions such as "ϕ is not false" and "ϕ is not true" in the object language:

$$
\begin{array}{rcll}
\circ\phi & \stackrel{def}{=} & ?\phi \vee \phi & \text{“}\phi\text{ \textit{is not false}”} \\
\bullet\phi & \stackrel{def}{=} & ?\phi \vee \neg\phi & \text{“}\phi\text{ \textit{is not true}”} \\
!\phi & \stackrel{def}{=} & \neg ?\phi & \text{“}\phi\text{ \textit{is not undefined}”}
\end{array}
$$

Furthermore, the sentence $\neg\bullet\phi$ makes the classical claim that ϕ is true, and $\neg\circ\phi$ makes the classical claim that ϕ is false. The fragment of Kleene's language discussed so far is unusual in that it has no tautologies; for example, the sentence $\phi \vee \neg\phi$ is undefined if ϕ is. The addition of the undefined operator produces three-valued analogues of classical tautologies; for example, the sentence $\phi \vee \bullet\phi$ and its equivalent $\phi \vee \neg\phi \vee ?\phi$ are tautologies.[1]

A drawback with Kleene's semantics is that the defined conditional operator, $\supset$, does not provide a satisfactory account of conditionality. In particular, if ϕ is true and ψ is undefined, then $\phi \supset \psi$ is undefined, rather than false. Moreover, in keeping with the resource-bounded interpretation, it is desirable to have a conditional which is true if the antecedent is not true, regardless of the truth value of the consequent. A conditional of this

[1] These operators have been investigated elsewhere. Mikhail Soutchanski observed that the undefined operator, ?, is a special case of the J_k operator of Rosser and Turquette [36], while the defined operator, !, is discussed by Avron [1], and the weak negation operator is discussed by van Benthem [47] and in the papers in [12].

kind can be thought of as expressing a constraint which must be met if the antecedent is true, but which can otherwise be ignored. It can be defined as follows:

$$\phi \rightarrow \psi \quad \stackrel{def}{=} \quad \neg\bullet\phi \supset \neg\bullet\psi$$

The new conditional is thus classically valued. It is false if the antecedent is true and the consequent is not, and true otherwise. Note in particular that the sentences $\phi \rightarrow \psi$ and $\bullet\psi \rightarrow \bullet\phi$ are equivalent, but that the sentences $\phi \rightarrow \psi$ and $\neg\psi \rightarrow \neg\phi$ are not.

despite a lack of expressiveness on

Kleene added his own classical operator. The sentence $\phi \equiv \psi$ states that ϕ and ψ are equivalent; that is, that they have the same truth value. This operator can be defined in the extended language as follows:

$$\phi \equiv \psi \quad \stackrel{def}{=} \quad (\neg\bullet\phi \wedge \neg\bullet\psi) \vee (\neg\circ\phi \wedge \neg\circ\psi) \vee (?\phi \wedge ?\psi)$$

Thus the sentences ϕ and ψ are equivalent if they are both true, or they are both false, or they are both undefined. For example the equivalence of a conditional sentence and its weak contrapositive, mentioned above, can now be stated in the object language: $(\phi \rightarrow \psi) \equiv (\bullet\psi \rightarrow \bullet\phi)$.

In order to represent time an explicit temporal index is added to each Kleene atom. However, rather than simply reifying time a predicative approach is adopted. Thus the temporal atom $r(u_1, \ldots, u_n)(t)$ states that the Kleene atom $r(u_1, \ldots, u_n)$ is true at the time point t. For example, we might write $At(O, L)(1)$ to represent the fact that Ollie is at location L at time 1. The temporal precedence operator, $<$, allows comparison of time points, thus $t < t'$ states that the time point t temporally precedes the time point t'. The semantics for time are as in standard temporal logic; see e.g. [35, 48]. Models now include a temporal frame $\langle \mathcal{T}, \mathcal{R}_\mathcal{T} \rangle$ where $\mathcal{R}_\mathcal{T}$ is a binary relation on $\mathcal{T}$ which is discrete and linear. Intuitively $\mathcal{T}$ is the set of time points and $\langle \mathcal{T}, \mathcal{R}_\mathcal{T} \rangle$ represents the flow of time. The introduction of time allows the truth value of Kleene literals to vary from time point to time point. For example, it may be the case that $At(O, L)(1)$ is true and $At(O, L)(2)$ is not. However, for simplicity, the denotation of terms is fixed across times. Accordingly the identity relation is atemporal. For example, we might write $L1 \neq L2$ to indicate that $L1$ and $L2$ are eternally distinct locations.

As in the Situation Calculus, the difference in aspect between facts and events is signified. Accordingly primitive event types are introduced as a third sort. Thus for example, $Move(x, l, l')$ might denote the movement of object x from location l to location l'. Event tokens are denoted by means of the occurs predicate. Thus a sentence of the form $Occ(e)(t)$ states that a token of an event of type e occurs at time t. For example, $Occ(Move(O, L1, L2))(1)$ states that Ollie moves from $L1$ to $L2$ at time 1.

In order to gain comparable expressivity with the Situation Calculus, *Kleene literals*, Kleene atoms and their negations, are added as a fourth sort. This higher-order quantification over literals is convenient, and can easily be replaced by treating Kleene literals as particulars and using a *Holds* predicate; as is done with fluents in the Situation Calculus. The final extension is a set of higher-order relations of the form $hr(e_1, \ldots, e_n, \ell_1, \ldots, \ell_m)(t)$; where $n + m > 0$, the e_i are event terms, and the ℓ_j are Kleene literals. A example is the *Occ* predicate discussed above. Further examples, such as *Aff* and *Qual*, are introduced in the next section.

2.2 The syntax and semantics of $\mathcal{TC}$

A formal account of the syntax and semantics $\mathcal{TC}$ is now given.

Definition 1. *The vocabulary of $\mathcal{TC}$ consists of the symbols:*

$$\neg, ?, \wedge, \forall, <, =, (,),$$

and the following mutually disjoint countable sets of symbols:

- D, T, E *(constants),*
- V_D, V_T, V_E, V_L *(variables),*
- F_D, F_T, F_E *(function symbols of arity $n \geq 1$),*
- R *(first-order relation symbols of arity $n \geq 0$), and*
- HR *(higher-order relation symbols of arity $n \geq 1$).*

Definition 2. *The terms of sorts material object, time point, event, and Kleene literal (the sets $term_D$, $term_T$, $term_E$, $term_L$ respectively) are defined as follows.*

- *If S is D or T then, $term_S$ is the minimal set such that*
  - $S \cup V_S \subseteq term_S$, *and*
  - *if f is an n-ary function symbol in F_S and $u_1, \ldots, u_n \in term_S$ then $f(u_1, \ldots, u_n) \in term_S$.*

- $term_E$ *is the minimal set such that*
  - $E \cup V_E \subseteq term_E$, *and*
  - *if* f *is an* n*-ary function symbol in* F_E *and* $u_1, \ldots, u_n \in term_D$ *then* $f(u_1, \ldots, u_n) \in term_E$.
- *Let* $A = \{r(u_1, \ldots, u_n) : \text{for } n\text{-ary } r \in R \text{ and } u_1, \ldots, u_n \in term_D\}$ *be the set of Kleene atoms and let* $L = A \cup \{\neg a : a \in A\}$ *be the set of Kleene literals, then* $term_L = L \cup V_L$.

Definition 3. $\mathcal{TC}$ *is the minimal set which satisfies the following conditions.*

- *If* $t, t' \in term_T$ *then* $t < t' \in \mathcal{TC}$.
- *If* $u_1, \ldots, u_n \in term_D$, $r \in R$ *is an* n*-ary first-order relation symbol, and* $t \in term_T$, *then* $r(u_1, \ldots, u_n)(t) \in \mathcal{TC}$.
- *If* $e_1, \ldots, e_n \in term_E$, $\ell_1, \ldots \ell_m \in L$, $hr \in HR$ *is an* $n+m$*-ary higher-order relation symbol, and* $t \in term_T$, *then* $hr(e_1, \ldots, e_n, \ell_1, \ldots \ell_m)(t) \in \mathcal{TC}$.
- *If* $\ell \in V_L$ *and* $t \in term_T$ *then* $\ell(t) \in \mathcal{TC}$.
- *If* S *is* D, T, E *or* L, *and* $u, u' \in term_S$, *then* $u = u' \in \mathcal{TC}$.
- *If* $\phi, \psi \in \mathcal{TC}$, *then* $\neg\phi \in \mathcal{TC}$, $?\phi \in \mathcal{TC}$, *and* $(\phi \wedge \psi) \in \mathcal{TC}$.
- *If* S *is* D, T, E *or* L, $v \in V_S$, *and* $\phi \in \mathcal{TC}$, *then* $\forall v \phi \in \mathcal{TC}$.

The members of $\mathcal{TC}$ *are called* formulas *(of* $\mathcal{TC}$*). Those formulas in which no variable occurs free are called* sentences *(of* $\mathcal{TC}$*).*

Definition 4. *A* model *for* $\mathcal{TC}$ *is a structure*

$$M = \langle \mathcal{D}, \mathcal{E}, \mathcal{T}, \mathcal{R}_\mathcal{T}, \mathcal{F}, \mathcal{R}, \mathcal{HR}, \mathcal{V} \rangle,$$

where:

- $\mathcal{D}$, $\mathcal{E}$ *and* $\mathcal{T}$ *are mutually disjoint non-empty sets,*
- $\mathcal{R}_\mathcal{T}$ *is a binary relation on* $\mathcal{T}$ *which is discrete and linear,*
- $\mathcal{F} = \langle \mathcal{F}_D, \mathcal{F}_T, \mathcal{F}_E \rangle$, *where* $\mathcal{F}_S$ *is a set of* n*-ary functions of type* $\mathcal{S}^n \to \mathcal{S}$, *for* $n \geq 1$ *and* $\langle S, \mathcal{S} \rangle \in \{\langle D, \mathcal{D} \rangle, \langle T, \mathcal{T} \rangle, \langle E, \mathcal{E} \rangle\}$,

- $\mathcal{R}$ *is a set of partial n-ary functions of type* $\mathcal{D}^n \rightarrow \{true, false\}$ *for* $n \geq 0$,

- $\mathcal{HR}$ *is a set of partial $n+m$-ary functions of type* $\mathcal{E}^n \times \mathcal{L}^m \rightarrow \{true, false\}$ *for* $n+m \geq 1$,

- $\mathcal{V} = \langle \mathcal{V}_D, \mathcal{V}_T, \mathcal{V}_E, \mathcal{V}_L, \mathcal{V}_{F_D}, \mathcal{V}_{F_T}, \mathcal{V}_{F_E}, \mathcal{V}_{MR}, \mathcal{V}_R, \rangle$ *is an interpretation function such that*

  - $\mathcal{V}_S : S \rightarrow \mathcal{S}$ *for* $\langle S, \mathcal{S}\rangle \in \{\langle D, \mathcal{D}\rangle, \langle T, \mathcal{T}\rangle, \langle E, \mathcal{E}\rangle\}$,
  - $\mathcal{V}_L : L \rightarrow L$ *is the identity function,*
  - $\mathcal{V}_{F_S} : F_S \rightarrow \mathcal{F}_S$ *for* $\langle S, \mathcal{S}\rangle \in \{\langle D, \mathcal{D}\rangle, \langle T, \mathcal{T}\rangle, \langle E, \mathcal{E}\rangle\}$,
  - $\mathcal{V}_R : R \times \mathcal{T} \rightarrow \mathcal{R}$, *and*
  - $\mathcal{V}_{HR} : HR \times \mathcal{T} \rightarrow \mathcal{HR}$.

Intuitively $\mathcal{D}$ is a set of material objects, $\mathcal{E}$ is a set of event types, $\mathcal{T}$ is a set of time points, and $\langle \mathcal{T}, \mathcal{R}_{\mathcal{T}}\rangle$ is a temporal frame which represents the flow of time. For each n-ary relation symbol r and time point t, $\mathcal{V}_R(r, t)$ is the partial characteristic function of the relation denoted by r at t. The definition of $\mathcal{V}$ allows the extension of relations to vary with time but, for simplicity, it keeps the denotation of constants and functions fixed across times.

Terms are interpreted in the standard way with the exception of literals, which are "interpreted" as themselves.

Definition 5. *A* variable assignment *for a* $\mathcal{TC}$*-model is a function* $g = \langle g_D, g_T, g_E, g_L\rangle$, *where* $g_L : V_L \rightarrow L$ *and otherwise* $g_S : V_S \rightarrow \mathcal{S}$. *The expression* $g \overset{v}{\approx} g'$ *means that that the variable assignments g and g' differ at most on what is assigned to the variable v.*

Definition 6. *Let M be a* $\mathcal{TC}$*-model with interpretation function* $\mathcal{V}$, *and let g be a variable assignment for M, then the* term evaluation function $\mathcal{V}_g$ *is defined as follows.*

- *For* $\langle S, \mathcal{S}\rangle \in \{\langle D, \mathcal{D}\rangle, \langle T, \mathcal{T}\rangle, \langle E, \mathcal{E}\rangle\}$

  - *if* $\tau \in S$ *then* $\mathcal{V}_g(\tau) = \mathcal{V}_S(\tau)$,
  - *if* $\tau \in V_S$ *then* $\mathcal{V}_g(\tau) = g_S(\tau)$,
  - *if* $\tau = f(u_1, \ldots, u_n) \in terms_S$ *then* $\mathcal{V}_g(\tau) = \mathcal{V}_{F_S}(f)(\mathcal{V}_g(u_1), \ldots, \mathcal{V}_g(u_n))$.

- *If* $\tau \in L$ *then* $\mathcal{V}_g(\tau) = \mathcal{V}_L(\tau) = \tau$*, and if* $\tau \in V_L$ *then* $\mathcal{V}_g(\tau) = g_L(\tau)$*.*

The truth and falsity of sentences can now be defined by means of the intermediary notions of the satisfaction and violation.

Definition 7. *Let* $M = \langle \mathcal{D}, \mathcal{E}, \mathcal{T}, \mathcal{R}_\mathcal{T}, \mathcal{F}, \mathcal{R}, \mathcal{HR}, \mathcal{V} \rangle$ *be a* $\mathcal{TC}$*-model,* g *be a variable assignment for* M*, and* ϕ *be a* $\mathcal{TC}$*-formula. Then* g satisfies ϕ *in* M *(written* $M, g \models \phi$*) or* violates ϕ *in* M *(written* $M, g \mathrel{=\!\!|} \phi$*) as follows:*

$$
\begin{array}{rcl}
M, g \models t < t' & \textit{iff} & (\mathcal{V}_g(t), \mathcal{V}_g(t')) \in \mathcal{R}_\mathcal{T} \\
M, g \mathrel{=\!\!|} t < t' & \textit{iff} & (\mathcal{V}_g(t), \mathcal{V}_g(t')) \notin \mathcal{R}_\mathcal{T} \\
M, g \models u = u' & \textit{iff} & \mathcal{V}_g(u) \ \textit{is} \ \mathcal{V}_g(u') \\
M, g \mathrel{=\!\!|} u = u' & \textit{iff} & \mathcal{V}_g(u) \ \textit{is not} \ \mathcal{V}_g(u') \\
M, g \models r(u_1, \ldots, u_n)(t) & \textit{iff} & \mathcal{R}(r, \mathcal{V}_g(t))(\mathcal{V}_g(u_1), \ldots, \mathcal{V}_g(u_n)) = \textit{true} \\
M, g \mathrel{=\!\!|} r(u_1, \ldots, u_n)(t) & \textit{iff} & \mathcal{R}(r, \mathcal{V}_g(t))(\mathcal{V}_g(u_1), \ldots, \mathcal{V}_g(u_n)) = \textit{false} \\
M, g \models hr(e_1, \ldots, e_n, & \textit{iff} & \mathcal{HR}(hr, \mathcal{V}_g(t))(\mathcal{V}_g(e_1), \ldots, \mathcal{V}_g(e_n), \\
\ell_1, \ldots, \ell_m)(t) & & \qquad \mathcal{V}_g(\ell_1), \ldots, \mathcal{V}_g(\ell_m)) = \textit{true} \\
M, g \mathrel{=\!\!|} hr(e_1, \ldots, e_n, & \textit{iff} & \mathcal{HR}(hr, \mathcal{V}_g(t))(\mathcal{V}_g(e_1), \ldots, \mathcal{V}_g(e_n), \\
\ell_1, \ldots, \ell_m)(t) & & \qquad \mathcal{V}_g(\ell_1), \ldots, \mathcal{V}_g(\ell_m)) = \textit{false} \\
M, g \models \neg\psi & \textit{iff} & M, g \mathrel{=\!\!|} \psi \\
M, g \mathrel{=\!\!|} \neg\psi & \textit{iff} & M, g \models \psi \\
M, g \models ?\psi & \textit{iff} & \textit{neither } M, g \models \psi \textit{ nor } M, g \mathrel{=\!\!|} \psi \\
M, g \mathrel{=\!\!|} ?\psi & \textit{iff} & \textit{either } M, g \models \psi \textit{ or } M, g \mathrel{=\!\!|} \psi \\
M, g \models \psi \wedge \chi & \textit{iff} & M, g \models \psi \textit{ and } M, g \models \chi \\
M, g \mathrel{=\!\!|} \psi \wedge \chi & \textit{iff} & M, g \mathrel{=\!\!|} \psi \textit{ or } M, g \mathrel{=\!\!|} \chi \\
M, g \models \forall v \psi & \textit{iff} & M, g' \models \psi \textit{ for all } g' \textit{ such that } g \stackrel{v}{\approx} g' \\
M, g \mathrel{=\!\!|} \forall v \psi & \textit{iff} & M, g' \mathrel{=\!\!|} \psi \textit{ for some } g' \textit{ such that } g \stackrel{v}{\approx} g'
\end{array}
$$

Definition 8. *A formula* ϕ *is* true *in a* $\mathcal{TC}$*-model* M *(written* $M \models \phi$*) if* $M, g \models \phi$ *for all variable assignments* g*. A formula* ϕ *is* false *in* M *(written* $M \mathrel{=\!\!|} \phi$*) if* $M, g \mathrel{=\!\!|} \phi$ *for all variable assignments* g*.*

It can be shown by induction on the structure of formulas (as in classical logic) that if a variable assignment satisfies a sentence in a model then every variable assignment satisfies the sentence in the model. Consequently either every variable assignment satisfies a sentence in a model or no variable assignment satisfies the sentence in the model. Similarly it can be shown that either every variable assignment violates a sentence in a model or no variable assignment violates the sentence in the model. Clearly also it follows from the clauses for negation that a sentence is false in a model if and only if its negation is true in the model. Thus:

Proposition 9. *Let M be a $\mathcal{TC}$-model and ϕ be a sentence of $\mathcal{TC}$. Then either $M \models \phi$ or $M \models \neg\phi$ or $M \models ?\phi$.*

Definition 10. *A $\mathcal{TC}$-model M is said to be a* model *of a sentence ϕ if $M \models \phi$. Similarly M is said to be a model of a set of sentences Θ (written $M \models \Theta$) if $M \models \phi$ for every $\phi \in \Theta$.*

3 Primary events

When it comes to the representation of events in Artificial Intelligence a good place to start, conceptually if not historically, is with the representation used in the planner STRIPS [13]. The states of the planner's world are represented by sets of classical first-order sentences. Each event type (each planning operator) is defined by an open atomic formula, representing the event name and its parameters, and three lists of formulas; a preconditions list, an additions list, and a deletions list. An event token is then an instantiation of the event type. For example, the unstacking of one block from another is is defined as the event type $Unstack(x, y)$ with preconditions and deletions list $\{On(x, y), Clear(x), HandEmpty\}$ and additions list $\{Holding(x), Clear(y)\}$. If the formulas in the preconditions list are true in a state then the event (the planning operator) is applicable, and applying it consists of adding the sentences in the additions list to the current state and removing the sentences in the deletions list from it. Events can thus be regarded as specialized deduction rules on states.

Events of this kind can be used as a basis for representing common sense causal reasoning. However, while the planning problem is to generate appropriate sequences of events in order to achieve a given goal or goals, the problem at hand is to represent the outcome of sequences of events, some of which may fail. Accordingly it is assumed that events occur regardless of whether their preconditions are true, and a distinction is drawn between the occurrence of an event and the success of the event.[2]

[2]This use of 'occurs' and 'succeeds' is common in the literature. It may sometimes appear to conflict with natural language usage, where it is often the case that to say an event occurs implies that it succeeds. For example to say that block A is unstacked from block B seems to imply that the unstack event succeeds. Allowance can easily be made for this reading 'occurs' as 'potentially occurs' and 'succeeds' as 'actually occurs'.

If an event occurs in a state and its preconditions are true in the state then the event succeeds and state is modified as before, otherwise the event fails and the state remains unchanged.[3]

In order to represent events of this kind in logic it is necessary to represent changes in state. When using $\mathcal{TC}$, states can be conceived of as being those facts which are true at each time point; thus, for example, $On(A, B)$ may be true at time 1 when an $Unstack(A, B)$ event occurs and false as a result at time 2. The STRIPS lists can be replaced by definitions giving the preconditions and the postconditions of event types. For example the definitions for the unstack event are as follows:[4]

$$\begin{aligned} Pre(Unstack(x, y))(t) &\equiv On(x, y)(t) \wedge Clear(x)(t) \wedge HandEmpty(t) \\ Post(Unstack(x, y))(t) &\equiv Holding(x)(t) \wedge Clear(y)(t) \wedge \\ &\quad\ \neg On(x, y)(t) \wedge \neg Clear(x)(t) \wedge \\ &\quad\ \neg HandEmpty(t) \end{aligned}$$

The effect of an event occurring can then be represented by an axiom of change:

$$Pre(e)(t) \wedge Occ(e)(t) \rightarrow Post(e)(t + 1) \tag{1}$$

Thus if event e occurs at time t and the preconditions of e are true at t then e succeeds and its postconditions are true at time $t + 1$.

3.1 Inertia

It is also necessary to represent what does not change from time point to time point. This is done by means of a common sense law of inertia which is very similar in spirit to the one proposed by McCarthy in his Situation Calculus formulation [28]. In keeping with the common sense intuition that nothing changes without reason, it seems appropriate to say that a fact persists (that a Kleene literal remains true) from time point to time point unless there is some reason for it not to do so. In order to do this the higher-order predicate *Aff* is used. Intuitively, for Kleene literal ℓ and time point t, $Aff(\ell)(t)$ states that the truth value of ℓ is affected at t; that is, that there

[3] From the planning perspective, applying an event whose preconditions are not true can be regarded as doing nothing; as performing a no-op, as "executing" a wait event, etc.

[4] As is customary, terms starting with upper case letters are used for constants and terms starting with lower case letters are used for variables. Customary abbreviations are also used. In particular universal quantifiers are omitted. Free variables should thus be understood as being bound by universal quantifiers with wide scope.

is reason to doubt that the truth value of ℓ persists beyond t. The common sense inertia axiom can now be stated as follows:[5]

$$\ell(t) \wedge \bullet \mathit{Aff}(\ell)(t) \rightarrow \ell(t+1) \tag{2}$$

Thus if ℓ is true at t and it is not true that ℓ is affected at t then ℓ remains true at $t+1$. This is the positive reading of the axiom. The contrapositive reading:

$$\ell(t) \wedge \bullet \ell(t+1) \rightarrow \mathit{Aff}(\ell)(t)$$

is, of course, true whenever the positive reading is. However the intention is that the axiom should be used positively whenever possible. The intention is that given $\ell(t)$, $?\mathit{Aff}(\ell)(t)$ should be assumed and the axiom used to conclude $\ell(t+1)$ if doing so is consistent.[6] Assumptions of this form will be referred to as *inertia assumptions.* As the inertia assumptions depend on the theory in which the axiom occurs, its interpretation will vary accordingly. In particular, it may behave non-monotonically. For example, $On(A,B)(2)$ should follow from the intended interpretation of the theory $\{On(A,B)(1),(2)\}$, but not from the intended interpretation of the extended theory $\{On(A,B)(1), \neg On(A,B)(2),(2)\}$.

Theories of the kind considered thus far—theories consisting of Kleene and occurs literals, together with the definitions of the preconditions and postconditions events and the change and inertia axioms—will be called *inertial theories.* The problem is thus to formalize the intended interpretation of inertial theories, to define a formal pragmatics for them. The literal, or semantic, interpretation of inertial theories is given by the semantics of $\mathcal{TC}$ which, in effect, define a semantic interpretation function, $[\![\cdot]\!]$, on arbitrary theories of $\mathcal{TC}$. For $\Theta \subseteq \mathcal{TC}$, $[\![\Theta]\!]$ is the set of $\mathcal{TC}$-models in which each sentence in Θ is true according to the semantics; that is, $[\![\Theta]\!] = \{M : M \text{ is a } \mathcal{TC}\text{-model and } M \models \Theta\}$. The problem is thus to define a pragmatic interpretation function, $[\cdot]$, such that $[\Theta]$ yields the intended interpretation of the inertial theory Θ.

The kind of causal reasoning discussed in this paper is ampliative; that is, the intended (or pragmatic) consequences of a theory include its literal (or semantic) consequences. Consequently, for any $\Theta \subseteq \mathcal{TC}$, it can be assumed that $[\Theta] \subseteq [\![\Theta]\!]$.The problem is thus to provide a definition of the pragmatic interpretation function, $[\cdot]$, which selects just those models from $[\![\Theta]\!]$ in which

[5] The unabbreviated axiom is, of course, $\forall \ell \forall t(\ell(t) \wedge \bullet \mathit{Aff}(\ell)(t) \rightarrow \ell(t+1))$.

[6] The not-true operator is used in the axiom in order to allow for its use if $\neg \mathit{Aff}(\ell)(t)$ is established independently.

Θ is interpreted as intended; or, equivalently, which removes just those models from $[\![\Theta]\!]$ in which Θ is not interpreted as intended.

McCarthy suggested, in effect, that for an inertial theory Θ, $[\Theta]$ should consist of those models of Θ in which changes are minimized. Consequently $[\Theta]$ can be defined to be the set of models of Θ in which affectations are minimized. Formally, for model M, let $M_{\mathcal{A}}$ be the set of affected sentences which are defined (which are true or false) in M; $M_{\mathcal{A}} = \{\mathit{Aff}(\ell)(t) : M \models\ !\mathit{Aff}(\ell)(t)\}$. A partial order, $\prec_{\mathcal{A}}$, is then defined on models by putting $M \prec_{\mathcal{A}} M'$ if and only if M and M' differ at most on the interpretation of Kleene literals and the *Occ* and *Aff* predicates and $M_{\mathcal{A}} \subset M'_{\mathcal{A}}$. Then a model M is said to be an $\mathcal{A}$-minimal model of a theory Θ if and only if $M \models \Theta$ and there is no model M' such that $M' \models \Theta$ and $M' \prec_{\mathcal{A}} M$, and $[\Theta]$ is the set of all $\mathcal{A}$-minimal models of Θ.

It seems that this pragmatics captures the intended interpretation of inertial theories. However while minimizing changes in this way eliminates some of the unintended models of inertial theories it does not eliminate all of them as the "canonical" example, due to Hanks and McDermott [15, 16], shows.

Example 11. (The Yale Shooting Problem) At time 1 a gun is loaded and pointed at Fred. Nothing relevant happens at time 2. At time 3 the gun is fired at Fred. If the gun is still loaded at time 3 then Fred is no longer alive as a result of the shooting. It seems natural to conclude that the gun is still loaded at time 3. The problem is to represent this common sense inference formally.

Let $\Theta_{11} = \{(1), (2), (3), (4), (5)\}$, where:

$$\mathit{Alive}(1) \wedge \mathit{Loaded}(1) \wedge \mathit{Occ}(\mathit{Shoot})(3) \tag{3}$$

$$\mathit{Pre}(\mathit{Shoot})(t) \equiv \mathit{Loaded}(t) \tag{4}$$

$$\mathit{Post}(\mathit{Shoot})(t) \equiv \neg\mathit{Alive}(t) \tag{5}$$

Then there are $\mathcal{A}$-minimal models of Θ_{11} in which the gun remains loaded as expected, and $\neg\mathit{Alive}(4)$ is true as a result. However there are also unintended $\mathcal{A}$-minimal models of Θ_{11} in which the gun is mysteriously unloaded during the pause at time 2 and so, unexpectedly, Fred survives.

Proof. The intended $\mathcal{A}$-minimal models of Θ_{11} correspond to those in which $?\mathit{Aff}(\mathit{Loaded})(1)$ and $?\mathit{Aff}(\mathit{Alive})(1)$ are assumed and the inertia axiom is used to infer $\mathit{Loaded}(2)$ and $\mathit{Alive}(2)$, and then $?\mathit{Aff}(\mathit{Loaded})(2)$ and $?\mathit{Aff}(\mathit{Alive})(2)$ are assumed and the inertia axiom is used to infer $\mathit{Loaded}(3)$ and $\mathit{Alive}(3)$. The shoot event thus succeeds and the change axiom, (1), gives $\neg Alive(4)$ as required. As $\neg Alive(4)$ is true in these models, it follows from the inertia axiom that $Aff(Alive)(3)$ is also true in them. However as this is the only affectation which is required to be true, and at least one affectation must be true in any model of Θ_{11},[7] it follows that the models described are indeed $\mathcal{A}$-minimal models of Θ_{11}.

The unintended $\mathcal{A}$-minimal models of Θ_{11} correspond to those in which $Alive(4)$ is assumed to be true. The truth of $Alive(4)$ is accounted for by the inertia axiom and the assumptions $?Aff(Alive)(3)$ and $?Aff(Alive)(2)$. But, as $Alive(4)$ is true, the shoot event which occurs at time 3 must fail. And, indeed, it follows from the change axiom that the precondition $Loaded(2)$ is not true. So it follows from the inertia axiom that $Aff(Loaded)(1)$ is true. Once again, this is the only affectation which is required, these models are $\mathcal{A}$-minimal models of Θ_{11}. $\square$

Hanks and McDermott observe that in this example the intended models are those which are obtained by reasoning forwards in time making inertia assumptions wherever possible, and that the unintended models are those which are obtained by reasoning backwards in time making inertia ssumptions wherever possible.

In keeping with this observation Shoham [42, 43] proposed that minimization be refined to chronological minimization and realised this in his classical epistemic temporal language TK. Chronological minimization can be realised in $\mathcal{TC}$ for inertial theories as follows. Let M and M' be models which differ at most on the interpretation of Kleene literals and the Occ and Aff predicates. Then M is chronologically less defined than M' (written $M \prec_{cld} M'$) if and only if there exists a time point t such that M and M' agree for any earlier time point, and at least one more Kleene or Occ or Aff literal is defined in M' at t. More formally, let $M_{\mathcal{R}}/t$ ($M_{\mathcal{O}}/t$, $M_{\mathcal{A}}/t$) be the set of Kleene literals (Occ literals, Aff literals) which are defined in M up to t:

$$\begin{aligned}
M_{\mathcal{R}}/t &= \{\ell(t') : t' \leq t \text{ and } M \models !\ell(t')\}, \\
M_{\mathcal{O}}/t &= \{Occ(e)(t') : t' \leq t \textit{ and } M \models !Occ(e)(t')\}, \\
M_{\mathcal{A}}/t &= \{Aff(\ell)(t') : t' \leq t \textit{ and } M \models !Aff(\ell)(t')\}.
\end{aligned}$$

Moreover, the union of these subsets will be denoted by the juxtaposition of their subscripts; thus, $M_{\mathcal{ROA}}/t$ denotes the set $M_{\mathcal{R}}/t \cup M_{\mathcal{O}}/t \cup M_{\mathcal{A}}/t$, etc.

[7]If $Loaded(3)$ is true in a model of Θ_{11} then, as noted above, $Aff(Alive)(3)$ is also true. While if $Loaded(3)$ is not true in the model then it follows from the inertia axiom that $Aff(Loaded)(2)$ is true.

Then $M \prec_{cld} M'$ if and only if M and M' differ at most on the interpretation of Kleene literals and the *Occ* and *Aff* predicates, and:

$$M_{\mathcal{ROA}}/t \subset M'_{\mathcal{ROA}}/t.$$

Then a model M is said to be a chronologically least defined model (a c.l.d. model) of a theory Θ if and only if $M \models \Theta$ and there is no model M' such that $M' \prec_{cld} M$ and $M' \models \Theta$, and $[\Theta]$ is the set of all c.l.d. models of Θ.

This refinement is sufficient to give the correct interpretation the inertia axiom in Hanks and McDermott's example:

Example 12. The c.l.d. models of Θ_{11} correspond to those $\mathcal{A}$-minimal models in which affectations are postponed for as long as possible, hence $\neg Alive(4)$ is true in all c.l.d. models of Θ_{11}.

Proof. Suppose that M is a c.l.d. model of Θ_{11}. Then it follows from the definition of c.l.d. model that $?Aff(Loaded)(1)$ and $?Aff(Alive)(1)$, and so it follows by the inertia axiom that $Loaded(2)$ and $Alive(2)$ are true in M. Similarly it follows that $Loaded(3)$ and $Alive(3)$ are true in M. $\square$

More generally, chronological minimization captures a necessary refinement of the interpretation of inertial theories. Given such a theory, the inertia assumptions should be made chronologically, with the effect that the axiom is applied chronologically. The rationale for this is as follows. Simply minimizing changes is inadequate as it does not take account of the relationship between change and inertia. In particular, it does not take into account the temporal directionality of change. Common sense has it that time is directed (by the before-after relation), that change results from events (that facts are affected by events and not conversely) and that change is temporally directed (that events precede their effects, or at least do not succeed them). This much is apparently clear in the syntax and proposed interpretation of inertial theories. However events only result in change if they succeed, their success depends on the presence of their preconditions, and the presence of the preconditions may depend on inertia. Consequently reasoning chronologically, "with the flow of events", gives a better representation of change and inertia than reasoning antichronologically, "against the flow of events". Hanks and McDermott's example illustrates this point nicely.[8]

[8] This justification of the principle of chronological minimization differs from those usually given. In particular, the justification for adopting the principle is not that it is desirable that nothing happens for as long as possible, or that "what you don't know won't hurt you"; the "ostrich principle". Chronological minimization has the effect of delaying change. But this is not the reason for adopting it.

3.2 Qualifications

Inertial theories thus provide the basis of a logical theory of STRIPS-type events. However, while the STRIPS view of events is simple and elegant, it fails to take account of many of the important aspects of common sense causal reasoning.

Perhaps the most fundamental limitation of this view of events is that it is deductive, rather than causal, in nature. If an event occurs and its preconditions are true then its success is certain and its postconditions are guaranteed to be true as a result. But in reality there are a potentially infinite number of reasons why an event may fail when its preconditions are true. For example, an unstack event may fail because the robot arm breaks, or the block is dropped, or the block is glued down, or the blocks world is perturbed by an experimenter or an earthquake, etc. It seems that in addition to stating the preconditions of an event we have to add an indefinite number of qualifications each of which states that some unusual disqualifying condition does not hold. McCarthy calls this the qualification problem and suggests another common sense principle: that we conclude by default that events succeed if their preconditions are true when they occur. In order to represent qualifications in $\mathcal{TC}$ the higher-order predicate $Qual$ is used. Intuitively $Qual(e)(t)$ states that event e is qualified at time t; that is, that there is reason to doubt that e succeeds at t. Then the deductive change axiom, (1), can be replaced by the following one:

$$Pre(e)(t) \wedge Occ(e)(t) \wedge \bullet Qual(e)(t) \rightarrow Post(e)(t+1) \tag{6}$$

Thus if event e occurs at time t and the preconditions of e are true at t and it is not true that e is qualified at t then the postconditions of e are true at $t+1$. The intention is to use this axiom positively whenever possible: given $Pre(e)(t)$ and $Occ(e)(t)$, $?Qual(e)(t)$ should be assumed and the axiom used to conclude $Post(e)(t+1)$, if doing so is consistent.[9] Assumptions of this form will be called *change assumptions*. Thus on the intended interpretation of the axiom events normally succeed if their preconditions are true when they occur.

The distinction between the occurrence of an event and its success is underlined by the following definition:

$$Succ(e)(t) \equiv Pre(e)(t) \wedge Occ(e)(t) \wedge \bullet Qual(e)(t). \tag{7}$$

[9] Once again, the not-true operator allows for the use of the axiom if $\neg Qual(e)(t)$ is established independently.

These two axioms together with the inertia axiom form the theory of primary events.

Definition 13. *The* theory of primary events*, Θ_P, consists of the inertia and change axioms: $\Theta_P = \{(2),(6),(7)\}$. Theories which contain Θ_P will be called* causal theories.

The problem now is to define an appropriate pragmatics for causal theories. Clearly, the success of an earlier event may lead to the failure of a later one, but the converse should not be the case. So, like the inertia axiom, the change axiom should be used chronologically, and consequently qualifications should be minimized chronologically. However as causal theories contain both the change and inertia axioms the possibility of conflicts between (the intended positive applications of instances of) these axioms arises.

Shoham avoids the issue by restricting his theories so that conflicts between common sense rules do not arise.[10] But this means that he cannot treat both inertia and qualifications in a single theory.

Suppose instead that the definition of chronological minimization is extended to include the *Qual* predicate. Thus let $M_{\mathcal{Q}}/t$ be the set of *Qual* literals which are defined in M up to t:

$$M_{\mathcal{Q}}/t \quad = \quad \{Qual(e)(t') : t' \leq t \text{ and } M \models !Qual(e)(t')\},$$

and now put $M \prec_{cld} M'$ if and only if M and M' differ at most on the interpretation of Kleene literals and the *Occ*, *Aff* and *Qual* predicates, and:[11]

$$M_{\mathcal{ROAQ}}/t \subset M'_{\mathcal{ROAQ}}/t.$$

Then the following simple example shows that chronological minimization is insufficient for causal theories.

Example 14. (The dodgy pickup) At time 1 block B is on the table and a pickup-B event occurs. Intuitively we think that the event succeeds, and that B is no longer on the table as a result.

Let $\Theta_{14} = \Theta_P \cup \{(8),(9),(10)\}$; where:

$$OnTable(B)(1) \wedge Occ(Pickup(B))(1) \tag{8}$$
$$Pre(Pickup(x))(t) \equiv OnTable(x)(t) \tag{9}$$
$$Post(Pickup(x))(t) \equiv \neg OnTable(x)(t) \tag{10}$$

---

[10] Such conflicts are forbidden by condition 7 of his definition of causal theories [43, p. 110].

[11] As before, juxtaposition of subscripts denotes the union of the sets, thus $M_{\mathcal{ROAQ}}/t = M_{\mathcal{R}}/t \cup M_{\mathcal{O}}/t \cup M_{\mathcal{A}}/t \cup M_{\mathcal{R}}/t$.

Then there are c.l.d. models of Θ_{14} in which $\neg OnTable(B)(2)$ is true as intended. However there are also unintended c.l.d. models of Θ_{14} in which B unexpectedly remains on the table.

Proof. In the intended models $?Qual(Pickup(B))(1)$ is true and so the pickup event succeeds; the change axiom is used positively giving $\neg OnTable(B)(2)$. As $\neg OnTable(B)(2)$ is true in these models, the positive use of the inertia axiom is blocked and it follows from the contrapositive use that $Aff(OnTable(B))(1)$.

In unintended models $?Aff(OnTable(B))(1)$ is true and so the positive use of the inertia axiom gives $OnTable(B)(2)$. Consequently the pickup event fails. The positive use of the change axiom is blocked and the contrapositive use gives $Qual(Pickup(B))(1)$. □

Where there is a conflict between the change axiom and the inertia axiom, it seems that the change axiom should take precedence. Consequently the pragmatics of causal theories should be refined so that selected models are those in which change is preferred to inertia.

Accordingly (simple) chronological minimization is refined to what might be called prioritized chronological minimization. The minimization is still chronological, however, at each time point, facts and events are minimized before qualifications which in turn are minimized before affectations. As the pragmatics is intended for causal theories, the selected models will be referred to as P-preferred models.

Definition 15. *Let M and M' be $\mathcal{TC}$-models which differ at most on the interpretation of Kleene literals and the Occ, Aff and Qual predicates. Then M is* P-preferred *to M' (written $M \prec_P M'$) if there exists a time point t such that:*

- $M_{\mathcal{RO}}/t \subset M'_{\mathcal{RO}}/t$, *or*
- $M_{\mathcal{RO}}/t = M'_{\mathcal{RO}}/t$ *and* $M_{\mathcal{Q}}/t \subset M'_{\mathcal{Q}}/t$, *or*
- $M_{\mathcal{ROQ}}/t = M'_{\mathcal{ROQ}}/t$ *and* $M_{\mathcal{A}}/t \subset M'_{\mathcal{A}}/t$.

Thus if M and M' are $\mathcal{TC}$-models which differ at most on the interpretation of Kleene literals and the *Occ*, *Aff* and *Qual* predicates, then $M \prec_P M'$ if there exists a time point t such that M and M' agree for any earlier time point and:

- at least one more Kleene or *Occ* literal is defined in M' at t, or
- M and M' agree on Kleene and *Occ* literals, and at least one more *Qual* literal is defined in M' at t, or

- M and M' agree on all Kleene, Occ, and $Qual$ literals, and at least one more Aff literal is defined in M' at t.

Thus $[\Theta]$ is now the set of P-preferred models of Θ and the pragmatic consequences of Θ are defined in the usual way:

Definition 16. *A model M is a P-*preferred model of *a sentence ϕ if $M \models \phi$ and there is no other model M' such that $M' \models \phi$ and $M' \prec_P M$. M is a P-preferred model of a set of sentences Θ if $M \models \Theta$ and there is no other model M' such that $M' \models \Theta$ and $M' \prec_P M$.*

Definition 17. *A causal theory Θ* predicts *a sentence ϕ, written $\Theta \mathrel{|\!\!\approx}_P \phi$, if ϕ is true in all P-preferred models of Θ.*

Clearly, as $\prec_P$ is a partial order, $\mathrel{|\!\!\approx}_P$ is a preferential entailment relation [43].

Example 18. Let Θ_{14} be as in Example 14. Then, as qualifications are minimized before affectations, $?Qual(Pickup(B))(1)$ is true in all P-preferred models of Θ_{14} and so the pickup event succeeds in all these models. Consequently $\Theta_{14} \mathrel{|\!\!\approx}_P \neg OnTable(B)(2)$.

Clearly if B is still on the table at time 2, say because the robot arm is broken, the pickup event should fail, and, indeed, $\Theta_{14} \cup \{OnTable(B)(2)\} \mathrel{|\!\!\approx}_P Qual(Pickup(B))(2)$. □

Two further properties of the pragmatics should be noticed. Firstly, at any time point, facts and events are minimized with equal priority. This is because there appears to be no intuitive basis for minimizing facts before events or vice versa. Secondly, minimizing facts and events before minimizing qualifications and affectations captures another important aspect of common sense causal reasoning: that speculating about the future does not alter the past or present. The effect of minimizing facts and events before qualifications and affectations at any time point is to fix the facts and events which are true at that time point (and consequently those which are true at all earlier time points) before making causal or inertial assumptions about the future. The beneficial consequences of this will emerge in the sequel. An immediate gain is that it is possible to reason inductively on the structure of time. This is illustrated by the following example which involves reasoning about inertia and the success of a temporally extended sequence of events.

Example 19. (The domino effect) At time 1 one hundred dominoes are arranged in a line such that toppling any of them results in the next one in the line toppling at time 2. If the first domino is toppled at time 1 we expect that this will initiate a sequence topplings which culminates with the toppling of the hundredth domino at time 100.

The functions d and l will be used to represent the dominoes and their locations respectively:

$$1 \leq x \leq 100 \rightarrow At(d_x, l_x)(1) \tag{11}$$

$$Occ(Topple(d_1, l_1))(1) \tag{12}$$

$$Pre(Topple(d_x, l_x))(t) \equiv At(d_x, l_x)(t) \tag{13}$$

$$\begin{aligned} &Post(Topple(d_x, l_x))(t) \equiv \\ &\qquad \neg At(d_x, l_x)(t) \wedge Occ(Topple(d_{x+1}, l_{x+1}))(t) \end{aligned} \tag{14}$$

Let $\Theta_{19} = \Theta_P \cup \{(11), \ldots, (14)\}$. Then as required, for all $n \leq 100$, $\Theta_{19} \approx_P Succ(Topple(d_n, l_n))(n)$.

Proof. By induction on time points (on positive integers).

For the base case let M be any P-preferred model of Θ_{19}. Then it follows from (11) that, for any m where $1 \leq m \leq 100$, $At(d_m, l_m)(1)$ is true in M. So in particular $At(d_1, l_1)(1)$ is true in M. Moreover it follows from (12) that $Occ(Topple(d_1, l_1))(1)$ is true in M. And it follows from the definition of P-preferred model that $?Qual(Topple(d_1, l_1))(1)$ is true in M. So it follows by (7) that $Succ(Topple(d_1, l_1))(1)$ is true in M.

For the induction step assume, for $n < 100$, that M is any P-preferred model of Θ_{19} in which $Succ(Topple(d_n, l_n))(n)$ is true and, for any m where $n \leq m \leq 100$, $At(d_m, l_m)(n)$ is true. Then it follows from the change axiom and (14) that $\neg At(d_n, l_n)(n+1)$ is true in M, and so it follows by the inertia axiom that $Aff(At(d_n, l_n))(n)$ is true in M. However, as $Topple(d_n, l_n)$ is the only event which succeeds at n and the only At-literal it affects at n is $At(d_n, l_n)$, it follows from the definition of P-preferred model that, for any m such that $n + 1 \leq m \leq 100$, $?Aff(At(d_m, l_m))(n)$ is true in M. So it follows by inertia that, for any m such that $n + 1 \leq m \leq 100$, $At(d_m, l_m)(n + 1)$ is true in M. In particular it follows that $At(d_{n+1}, l_{n+1})(n + 1)$ is true in M. Moreover, as $Succ(Topple(d_n, l_n))(n)$ is true in M, it follows from the change axiom and (14) that $Occ(Topple(d_{n+1}, l_{n+1}))(n + 1)$ is true in M. And it follows from the definition of P-preferred model that $?Qual(Topple(d_{n+1}, l_{n+1}))(n + 1)$ is true in M. So it follows from the success axiom, (7), that $Succ(Topple(d_{n+1}, l_{n+1}))(n + 1)$ is true in M. $\square$

3.3 Simultaneous events

A second limitation of the STRIPS view of events is that it it relies on what is often called the STRIPS assumption, the assumption that only one event occurs at a time. Removing this assumption in STRIPS can lead to conflicts between events, resulting in incongruous states and even in contradictions.[12] However the causal nature of events in causal theories which results from the treatment of qualifications provides a basis for the representation of simultaneous events, essentially because potential conflicts between events give rise to additional qualifications.

In illustrating this, *domain axioms* and *uniqueness of names axioms* are introduced. Domain axioms are sentences of $\mathcal{TC}$ which are *synchronic*, or temporally undirected, in that all temporal references in them are to the same time; unlike the change and inertia axioms which are *diachronic*, or temporally directed, as they refer "across time" in their antecedents and consequents. Uniqueness of names axioms were introduced by Baker [2], and are simply convenient abbreviations for conjunctions of inequalities. Thus, for terms $u_1, \ldots, u_n$:

$$UNA[u_1, \ldots, u_n] \stackrel{def}{=} \bigwedge u_i \neq u_j \text{ for all } 1 \leq i < j \leq n$$

Example 20. At time 1 Ollie is at location $L1$, Stan is at location $L3$ and both attempt to move to location $L2$:

$$At(O, L1)(1) \wedge At(S, L3)(1) \tag{15}$$
$$Occ(Move(O, L1, L2))(1) \wedge Occ(Move(S, L3, L2))(1) \tag{16}$$
$$Pre(Move(x, l, l'))(t) \equiv At(x, l)(t) \tag{17}$$
$$Post(Move(x, l, l'))(t) \equiv At(x, l')(t) \wedge \neg At(x, l)(t) \tag{18}$$
$$At(x, l)(t) \wedge x \neq y \rightarrow \neg At(y, l)(t) \tag{19}$$
$$UNA[S, O, L1, L2, L3] \tag{20}$$

Thus (19) is a domain axiom which states that two different objects cannot be at the same location at the same time, and (20) states that the names S, O, $L1$, $L2$ and $L3$ refer to different objects.

Let $\Theta_{20} = \Theta_P \cup \{(15), \ldots, (20)\}$. Then there are P-preferred models of Θ_{20} in which Ollie succeeds in getting to $L2$ at time 2 and Stan remains

[12] For example Nilsson [31, p. 339] defines stack and unstack operators which can be performed simultaneously with different hands. However a conflict arises if two hands attempt to perform simultaneous operations on the same block. Thus if $Unstack(A, B, H1)$ and $Unstack(A, B, H2)$ occur simultaneously and the preconditions of each event are true, the incongruous result is that both hands are holding A. If $Stack(A, B, H1)$ then occurs, an incongruous state arises in which A is clear and $H2$ is holding A. A contradiction now arises if $Stack(A, A, H2)$ occurs, as $Clear(A)$ has to be both added and deleted as a result.

at $L3$. There are also P-preferred models of Θ_{20} in which Stan succeeds in getting to $L2$ at time 2, and Ollie remains at $L1$. Consequently neither $\Theta_{20} \not\approx_P At(O, L2)(2)$ nor $\Theta_{20} \not\approx_P At(S, L2)(2)$.

Proof. Let M be a P-preferred model of Θ_{20} in which $?Qual(Move(O, L1, L2))(1)$ is true. Then it follows, by the success axiom and (15)-(17), that $Succ(Move(O, L1, L2))(2)$ is true in M. And it follows by (15)-(18) and the change axiom that $At(O, L2)(2)$ is true in M. Consequently it follows from (19) and (20) that $\neg At(S, L2)(2)$ is true in M. So it follows by the change axiom that $Qual(Move(S, L3, L2))(1)$ is true in M. So, as M is a P-preferred model of Θ_{20}, it follows by inertia that $At(Stan, L3)(2)$ is true in M.

Analogous reasoning shows that there are also P-preferred models of Θ_{20} in which $At(S, L2)(2)$ and $At(O, L1)(2)$ are true. □

Example 21. Example 20 can be elaborated by stating that when conflicting move events occur Ollie succeeds, say because he is bigger:

$$Occ(Move(O, l, l'))(t) \wedge Occ(Move(S, l'', l'))(t) \rightarrow Qual(Move(S, l'', l'))(t) \tag{21}$$

Then, as Stan's attempt to move fails in all models of $\Theta_{20} \cup \{(21)\}$, $\Theta_{20} \cup \{(21)\} \approx_P At(O, L2)(2)$. □

In this example axiom (21) has the desired effect. If Stan and Ollie both attempt to move to the same location, then Stan fails and Ollie succeeds. So the conflict is resolved in Ollie's favour. However if there is some independent reason why Ollie fails, say because he slips, then the axiom has the unfortunate consequence that Stan also fails. Clearly what is needed is some means of saying that Ollie's success is preferred to Stan's without it being the case that Stan's failure is prohibited should Ollie fail for some independent reason.

In order to be able to do so, a binary $Qual$ relation will be used. Thus $Qual(e, e')(t)$ should be thought of as stating that if events e and e' occur at time t then e normally qualifies e' at t, that the success of event e is more likely than the success of event e', that the success of e is preferred to the success of e'. Thus if there is no independent reason reason why e should fail (if there is no reason other than the occurrence of e'), then e should succeed and e' should fail. However, if e does fail for some independent reason, then the occurrence of e should not, of itself, prevent e' from succeeding. Conflict resolution axioms such as (21) can thus be replaced by statements of preference regarding qualifications; for example, (21) can be replaced by the statement $Qual(Move(O,l,l'), Move(S,l'',l'))(t)$. An instance of a qualification preference satement is said to be applicable at time t if both of the events referred to therein occur at t. If a particular preference statement is applicable at t, then one of the events involved must be qualified:

$$Occ(e)(t) \wedge Occ(e')(t) \wedge Qual(e,e')(t) \rightarrow Qual(e)(t) \vee Qual(e')(t) \tag{22}$$

Qualification preferences are also required to be asymmetric:

$$Qual(e,e')(t) \rightarrow Qual(e',e)(t) \tag{23}$$

Consequently any causal theory containing qualification preference satements should also contain the axioms (22) and (23), which will be referred to as the theory of qualification preferences, Θ_Q.

Definition 22. *The* theory of qualification preferences, *Θ_Q, consists of the axioms* (22) *and* (23)*, and the* theory of primary events with qualification preferences, *Θ_{PQ}, is $\Theta_P \cup \Theta_Q$.*

Of course, the pragmatics of causal theories containing Θ_Q needs further refinement in order to obtain the intended interpretation of tits qualification preference statements. In particular, at each time point, it is necessary to provide a means of choosing between qualifications in accordance with the qualification preferences expressed in the theory. Thus if $M_Q/t \not\subseteq M'_Q/t$ then M may still be preferred to M' on the basis of the qualification preferences which are applicable at time t, written $M_Q/t \prec_Q M'_Q/t$. Informally, $M_Q/t \prec_Q M'_Q/t$ if either fewer qualification preferences are applicable in M than in M', or the same qualification preferences are applicable in M and M' at t, M and M' agree on all qualifications at t which are not referred to in the qualification preferences,[13] and M better satisfies the qualification preferences than M' does.

Formally, for model M, let $M_Q(t)$ be the set of all events which are qualified in M at time t:

$$M_Q(t) = \{e : M \models Qual(e)(t)\},$$

let:

$$P_M(t) = \{Qual(e,e')(t) : M \models Occ(e)(t) \wedge Occ(e')(t) \wedge Qual(e,e')(t)\}$$

[13]By supposition M and M' agree on all earlier qualifications; that is, $M_Q/t-1 = M'_Q/t-1$.

be the set of qualification preferences which are applicable in M at time t, and let:

$$E_M(t) = \{e : \mathit{Qual}(e, e')(t) \in P_M(t) \vee \mathit{Qual}(e', e)(t) \in P_M(t)\}.$$

be the set events referred to in the qualification preferences. As the subsequent definition compares models in which the same qualification preferences are applicable, and does so only a time t, references to particular models and to t will be suppressed; thus $P_M(t)$ will be abbreviated to P, etc. The following abbreviation will also be useful:

$$e \prec_q e' \quad \text{iff} \quad \mathit{Qual}(e, e')(t) \in P.$$

Thus if $e \prec_q e'$, then the non-qualification of e is preferred to the non-qualification of e' (at t).

Now, for models M and M' (which agree on P at t) M is at least as preferred as M' regarding the qualification of event e (at time t) as follows:

$$\begin{aligned} M \preceq_e M' \quad \text{iff} \quad & e \in M'_Q \text{ if } e \in M_Q, \text{ and} \\ & \text{for every } e' \text{ such that } e' \prec_q e, e' \in M_Q \text{ iff } e' \in M'_Q. \end{aligned}$$

Thus M is at least as preferred as M' regarding the qualification of e if and only if M and M' agree on the qualification status of all events whose non-qualification is preferred to that of e, and e is qualified in M' if e is qualified in M. So M is preferred to M' regarding the qualification status of e if and only if M is at least as preferred as M on e, and e is qualified in M' but not in M:

$$M \prec_e M' \quad \text{iff} \quad M \preceq_e M' \text{ and } e \notin M_Q \text{ and } e \in M'_Q.$$

Now M satisfies the qualification preferences at least as well as M' does (at time t) if and only if, for every event e which is referred to in an applicable qualification preference, either M is at least as preferred as M' regarding the qualification of e, or there is an event e' whose non-qualification is preferred to that of e and M is preferred to M' on e':

$$\begin{aligned} M \preceq_Q M' \quad \text{iff} \quad & \text{for any } e \in E, \text{ either } M \preceq_e M', \mathit{or} \\ & \text{there is an } e' \text{ such that } e' \prec_q e \text{ and } M \prec_{e'} M'. \end{aligned}$$

And M satisfies the qualification preferences better than M' does (at time t) if and only if M satisfies every relevant event at least as well as M' does and M satisfies at least one event better than M' does:

$$\begin{aligned} M \prec_Q M' \quad \text{iff} \quad & M \preceq_Q M', \\ & \text{and there is an } e \in E \text{ such that } M \prec_e M'. \end{aligned}$$

The preference relation $\prec_Q$ is easily shown to be transitive. Suppose that $M \prec_Q M'$ and that $M' \prec_Q M''$. Then, as $M \preceq_Q M'$ and $M' \preceq_Q M''$, it follows that $M \preceq_Q M''$. And, as $M \prec_Q M'$ and $M' \preceq_Q M''$, there is an event $e \in E$ such that $M \prec_e M'$ and $M' \preceq_e M''$, so $M \prec_e M''$.

The full definition of the $\prec_{\mathcal{Q}}$ relation can now be given:

$$\begin{array}{lll} M_{\mathcal{Q}}/t \prec_{\mathcal{Q}} M'_{\mathcal{Q}}/t & \text{iff} & \text{either } P_M(t) \subset P_{M'}(t), \text{ or} \\ & & P_M(t) = P_{M'}(t), \\ & & M_{\mathcal{Q}}(t) - E_M(t) = M'_{\mathcal{Q}}(t) - E_{M'}(t), \\ & & \text{and } M \prec_Q M'. \end{array}$$

Finally, the definition of a P-preferred model is extended to that of a PQ-preferred model by refining the choice at the qualification stage:

Definition 23. *Let M and M' be $\mathcal{TC}$-models which differ at most on the interpretation of Kleene literals and the Occ, Aff and Qual predicates. Then M is PQ-*preferred *to M' (written $M \prec_{PQ} M'$) if there exists a time point t such that:*

- *$M_{\mathcal{RO}}/t \subset M'_{\mathcal{RO}}/t$, or*
- *$M_{\mathcal{RO}}/t = M'_{\mathcal{RO}}/t$ and either $M_{\mathcal{Q}}/t \subset M'_{\mathcal{Q}}/t$ or $M_{\mathcal{Q}}/t \prec_{\mathcal{Q}} M'_{\mathcal{Q}}/t$, or*
- *$M_{\mathcal{ROQ}}/t = M'_{\mathcal{ROQ}}/t$ and $M_{\mathcal{A}}/t \subset M'_{\mathcal{A}}/t$.*

Clearly $\prec_{PQ}$ is transitive. Clearly also PQ-preference is equivalent to P-preference in the absence of applicable qualification preferences. The notions of PQ-preferred model and PQ-entailment follow those for P.

Definition 24. *A model M is a PQ-preferred model of a sentence ϕ if $M \models \phi$ and there is no other model M' such that $M' \models \phi$ and $M' \prec_{PQ} M$. M is a PQ-preferred model of a set of sentences Θ if $M \models \Theta$ and there is no other model M' such that $M' \models \Theta$ and $M' \prec_{PQ} M$.*

Definition 25. *A theory Θ which contains Θ_{PQ}* predicts *a sentence ϕ, written $\Theta \mathrel{|\!\!\approx}_{PQ} \phi$, if ϕ is true in all PQ-preferred models of Θ.*

Clearly, $\mathrel{|\!\!\approx}_{PQ}$ is a preferential entailment relation.

Example 26. Suppose that a qualification preference statement is added to the theory of Example 20 stating that if Stan and Ollie attempt to move to the same location simultaneously then Ollie's success is preferred to Stan's. Suppose, that is, that $\Theta_{26} = \Theta_{26} \cup \{Qual(Move(O,l,l'), Move(S,l'',l'))(t)\}$.

Then Θ_{26} predicts that Ollie succeeds and Stan fails, $\Theta_{26} \approx_{PQ} At(O,L2)(2) \wedge At(S,L3)(2)$.

However if Ollie fails for some independent reason, if he slips say, then Stan should succeed in moving to the new location. Thus let $\Theta'_{26} = \Theta_{26} \cup \{(24),(25)\}$, where:

$$Occ(Slips(O))(1) \tag{24}$$
$$Occ(Move(x,l,l'))(t) \wedge Occ(Slips(x))(t) \rightarrow Qual(Move(x,l,l'))(t) \tag{25}$$

Then, as desired, $\Theta'_{26} \approx_{PQ} At(O,L1)(2) \wedge At(S,L2)(2)$.

Proof. Consider two models M and M' which are P-preferred models of Θ_{26} and which agree up to the qualification stage of the definition of PQ-preferred model. Suppose further that both models agree on the minimal set of qualification statements at time 1, that:

$$P_M(1) = P_{M'}(1) = \{Qual(Move(O,L1,L2), Move(S,L3,L2))(1)\}.$$

but that the models disagree on which of the two events referred to in the preference statement succeeds, that:

$$M_Q(1) = \{Move(S,L3,L2)\} \text{ and } M'_Q(1) = \{Move(O,L1,L2)\}.$$

Then $M \prec_{Move(O,L1,L2)} M'$ and $M \preceq_Q M'$. So $M \prec_Q M'$, and consequently $M \prec_{PQ} M'$.

In any model of Θ'_{26}, $Qual(Move(O,L1,L2))(1)$ is true by (24) and (25). So in any preferred PQ-model of Θ'_{26} (or, indeed, in any preferred P-model of Θ'_{26}), $?Qual(Move(S,L3,L2))(1)$ and $Succ(Move(S,L3,l2))(1)$ are true. $\Box$

The next example shows that the definition gives appropriate results when the qualification preferences are not transitive.

Example 27. (Three stooges) Suppose that initially Ollie is at location $L1$, Stan is at location $L2$, and Charlie is at location $L3$. If Stan and Ollie attempt to move to location $L4$ simultaneously, then Ollie's success is preferred to Stan's. Similarly if Stan and Charlie attempt the move simultaneously, then Stan's success is preferred to Charlie's. While if Ollie and Charlie attempt the move simultaneously, then Charlie's success is preferred to Ollie's.

This situation can be represented by the causal theory $\Theta_{27} = \Theta_{PQ} \cup \{(17),(18),(19),(26),\ldots,(30)\}$, where:

$$
\begin{aligned}
&At(O,L1)(1) \wedge At(S,L2)(1) \wedge At(C,L3)(1) && (26)\\
&Qual(Move(O,l,l'),Move(S,l,l'))(t) && (27)\\
&Qual(Move(S,l,l'),Move(C,l,l'))(t) && (28)\\
&Qual(Move(C,l,l'),Move(O,l,l'))(t) && (29)\\
&UNA[S,O,C,L1,L2,L3,L4] && (30)
\end{aligned}
$$

Now if, additionally, two of the move events occur simultaneously the preferred event succeeds; as in the previous example. However if all three move events occur simultaneously, the result is chaos; it is not possible to predict where any of the agents ends up.

Proof. Suppose that all three move events occur at time 1. Suppose, that is, that $\Theta_{27'} = \Theta_{27} \cup \{(31)\}$; where:

$$
\begin{aligned}
&Occ(Move(O,L1,L4))(1) \wedge\\
&Occ(Move(S,L2,L4))(1) \wedge\\
&Occ(Move(C,L3,L4))(1) \qquad (31)
\end{aligned}
$$

Consider three models M, M', and M'' which are P-preferred models of $\Theta_{27'}$, which agree up to the qualification stage of the definition of PQ-preferred model, but which disagree in that:

$$
\begin{aligned}
M_Q(1) &= \{Move(S,L2,L4), Move(C,L3,L4)\},\\
M'_Q(1) &= \{Move(O,L1,L4), Move(C,L3,L4)\},\\
M''_Q(1) &= \{Move(O,L1,L4), Move(S,L2,L4)\}.
\end{aligned}
$$

Suppose further that all three models agree on the minimal set of applicable qualification preferences at time 1, that:

$$
\begin{aligned}
P(1) = \{&Qual(Move(O,L1,L4),Move(S,L2,L4))(1),\\
&Qual(Move(S,L2,L4),Move(C,L3,L4))(1),\\
&Qual(Move(C,L3,L4),Move(O,L1,L4))(1)\}.
\end{aligned}
$$

Then, in order to show that $N \prec_Q N'$ (where N and N' are any two of the three models M, M' and M''), it is necessary to show that $N \preceq_Q N'$; that is, that, for any event $e \in E(1)$, either $N \preceq_e N'$ or there is an $e' \in E(1)$ such that $e' \prec_q e$ and $N \prec_{e'} N'$. However, it is not possible to show this for a single such event. As the preferences on the events in $E(1)$ are cyclical, it is only possible to show that $N \preceq_e N'$ if $N_Q/t \subseteq N'_Q/t$. But, in each case, $N_Q/t \not\subseteq N'_Q/t$. So, in each case, $N \not\prec_Q N'$. Each of the three events is preferred to its preference-companion in one model, but is qualified by an independent event in another model. The consideration of qualification preferences thus does not discriminate between the models and so $\Theta_{27'}$ is silent about the outcome of the three move events and the whereabouts of the three stooges at time 2. $\square$

3.4 Ramifications

Causal theories thus seem to provide a suitable basis for representing change involving inertia and qualifications in the presence of simultaneous events. However another unrealistic restriction has been inherited from the STRIPS representation of events. If a STRIPS event occurs successfully its effects are invariant; the facts in its add list are always added to the current state and the facts in its delete list are always removed from it. However in reality the effects of a successful event may vary according to the context in which it occurs. In the literature these effects are commonly referred to as ramifications.[14] Ramifications pose a problem for causal theories as the following example, due to Baker and Lifschitz, shows.[15]

Example 28. Initially Ollie is holding a block at location $L1$ and moves to location $L2$. Intuitively it seems that the move event should succeed and that, as Ollie is holding the block when he moves, it should move with him.

It seems that this scenario can be represented by the causal theory $\Theta_{28} = \Theta_P \cup \{(17), (18), (32), \ldots, (36)\}$; where:

$$At(O, L1)(1) \wedge Holds(O, B)(1) \tag{32}$$

$$Occ(Move(O, L1, L2))(1) \tag{33}$$

$$At(x, l)(t) \wedge l \neq l' \rightarrow \neg At(x, l')(t) \tag{34}$$

$$At(x, l)(t) \wedge Holds(x, y)(t) \rightarrow At(y, l)(t) \tag{35}$$

$$UNA[O, B, L1, L2] \tag{36}$$

[14] Variation also arises in the subsequent effects of events in a temporally extended chain of events. Thus in the dominoes example (Example 19), toppling the first domino results in the remaining dominoes in the line toppling in sequence at successive points in time. However if one of the dominoes is prevented from toppling (by being removed, held up, glued down, etc.) the process is interrupted and so the effects of the initial toppling event are changed. Variations of this sort are easily represented in causal theories.

[15] The example is essentially the same as Baker's ice cream example [3] and is analogous to Lifschitz's lamp circuit example [22].

However, contrary to intention, $\Theta_{28} \not\approx_P At(B, L2)(2)$.

Proof. As change is preferred to inertia the move event succeeds in all P-preferred models of Θ_{28} and so $At(O, L2)(2)$ is true in all of them.

In the intended P-preferred models of Θ_{28} reasoning about inertia (assuming $?\mathit{Aff}(Holds(O, B))(1)$ and using (32) and the inertia axiom) gives $Holds(O, B)(2)$ so axiom (35) gives $At(B, L2)(2)$.

However, $At(B, L1)(1)$ is true by (32) and (35), so there are also unintended P-preferred models of Θ_{28} in which reasoning about inertia (assuming $?\mathit{Aff}(At(B, L1))(1)$ and using the inertia axiom) gives $At(B, L1)(2)$. In these models $\neg At(B, L2)(2)$ is true, by axioms(34) and (36). So axiom (35) gives $\bullet Holds(O, B)(2)$. $\Box$

In this example the movement of the block is a ramification as it is an additional effect of Ollie's movement which only arises because he is holding the block when he moves. But the attempt to represent it by means of axiom (35) fails. The problem is that, unlike domain axioms such as (19) and (34), axiom (35) only has the desired effect if it is interpreted appropriately. Clearly the intention is that it should be interpreted positively; and, indeed, the intended P-preferred models of Θ_{28} are those where this is the case. But the pragmatics of causal theories is designed to yield the appropriate interpretation of the diachronic change and inertia axioms and has no effect on the interpretation of synchronic domain axioms. Consequently the unintended contrapositive interpretation of (35) is not eliminated; and this results in the unintended P-preferred models of Θ_{28}.

One might attempt to solve this problem by introducing a "causally directed" conditional which could only be interpreted positively. However even if this could be done the example raises a more general problem. The intended positive interpretation of axiom (35) depends on appropriate reasoning about inertia, and in particular on the appropriate use of the inertia axiom; it is necessary to reason that Ollie keeps hold of the block, rather than reasoning that it remains stationary. But it seems paradoxical that reasoning about change should be based on reasoning about inertia, and consequently it seems methodologically undesirable to use the inertia axiom, in conjunction with domain axioms, to represent reasoning about change.[16]

What then accounts for our intuitive understanding of the causal story of Example 28? Why do we think that when Ollie moves, the block moves with him? It seems that the reason why we do think this is that we are told

[16] More precisely, it seems undesirable to use the inertia axiom to reason about the effects of events. It may properly be used to establish the context in which they occur; for example to reason that the gun remains loaded in Example 12.

that he is holding the block when he moves and we are not told that he puts it down. Consequently we discount the possibility of the block remaining at $L1$ because an additional event would be required in order to account for this. However the movement of the block is itself an additional event which has the additional effect that the block is at $L2$. The missing causal element in this example is thus an event, and the choice between a move event and a putdown event seems clear. However the block's moving differs from Ollie's moving in that the block moves only because Ollie moves and only because he is holding the block when he does so. In order to reflect this difference, a distinction is drawn between primary and secondary events.

4 Secondary events

Primary events are events of the kind considered so far. By contrast, secondary events are additional events which are invoked by primary events in appropriate contexts and which, if they succeed, produce the additional context-dependent effects of the primary events which invoke them. Thus, in Example 28, Ollie's moving is a primary event which, as it occurs in a context in which he is holding the block, invokes the secondary event of the block's moving and, as both events succeed, the block's moving results in the additional context-dependent effect of it changing location.[17]

Secondary events occur simultaneously with the primary events which invoke them, and occur only because they are invoked; thus in the example the move-Ollie event and the move-block event occur simultaneously and the latter only occurs because it is invoked by the former. A secondary event may be invoked by more than one primary event; for example, if Stan and Ollie are both holding the block and they move in unison then the move-Stan and move-Ollie events both invoke the move-block event.

Moreover the success of secondary events depends on the success of primary events. A secondary event can only succeed if it is invoked by a successful primary event, but the converse need not be the case. Thus in

[17] Primary and secondary events are so-called by loose analogy with the philosophical distinction between primary and secondary qualities. This distinction "is especially associated with the 17th-century rise of modern science, with its recognition that the fundamental explanatory properties of things are not the qualities that perception most immediately concerns. These latter are the secondary qualities, or immediate sensory qualities, including colour, taste, smell, felt warmth or texture, and sound. The primary properties are less tied to the deliverance of one sense, and include the size, shape and motion of objects" [11].

The term 'invoke' is being used here in a way that is common in Computer Science, where one program (procedure, process, ...) is said to invoke another.

Example 28 if the move-Ollie event fails, the move-block event fails also. But if Stan and Ollie are both holding the block and only Ollie moves then the move-Ollie event may succeed and the move-block event may fail.

The distinction between primary and secondary events is relative. If an event e is invoked by another event e', then e is considered to be secondary, otherwise e is considered to be primary. In other circumstances, e might occur independently of e' as a primary event in its own right.[18] Moreover, a secondary event may in turn have context-dependent effects, so there should be tertiary, and in general n-ary, events.[19]

Events can now be thought of as structured objects, and their effects can be thought of as including not just their direct effects (their postconditions and the consequences of these given the domain axioms) but also their ramifications (the direct effects of any events invoked by them). The effects of an event can thus be seen to vary according to the context in which it occurs.

In causal theories the invocation of events is represented by means of the higher-order relation *Inv* and invocation axioms. Intuitively $Inv(e, e')(t)$ states that event e invokes event e' at time t. Invocation axioms can thus be thought of as having the following form:

$$Occ(e)(t) \wedge \Phi(e, e')(t) \rightarrow Inv(e, e')(t) \tag{37}$$

where $\Phi(e, e')$ is a formula which distinguishes those contexts in which e invokes e'. In order to represent the properties of secondary events the invocation relation is required to satisfy the following axioms:

$$\neg Inv(e, e)(t) \tag{38}$$

$$Inv(e, e')(t) \wedge Inv(e', e'')(t) \rightarrow Inv(e, e'')(t) \tag{39}$$

$$Inv(e, e')(t) \rightarrow Occ(e)(t) \wedge Occ(e')(t) \tag{40}$$

$$\begin{aligned} & Inv(e, e')(t) \wedge Succ(e')(t) \rightarrow \\ & \qquad Succ(e)(t) \vee \exists e''(Inv(e'', e')(t) \wedge Succ(e'')(t)) \end{aligned} \tag{41}$$

Thus axioms (38) and (39) state that invocation is irreflexive and transitive; and hence asymmetric. Axiom (40) requires that both the invoking and

[18]The analogy with primary and secondary qualities is thus imperfect. The intended points of similarity are that secondary events are dependent on primary events, just as secondary qualities are dependent on primary qualities, and that secondary events are context dependent, just as secondary properties are; e.g. an object's colour may differ to different observers and in different lighting conditions.

[19]Similarly, tertiary qualities are "Qualities or powers a thing has in virtue of its secondary qualities, just as they exist in virtue of its primary qualities: the flower is attractive to the butterfly because of its colour, or the wine is expensive because of its taste" [11].

invoked events occur. Finally Axiom (41) ensures that an invoked event succeeds only if it is invoked by a successful event. Note, however, that it may be the case the invoking event succeeds and the invoked event fails.[20]

The invocation axioms form the theory of secondary events.

Definition 29. *The* theory of secondary events, Θ_S, *consists of the invocation axioms,* $\Theta_S = \{(38), \ldots, (41)\}$. *The* theory of primary and secondary events *is the combination of the two theories,* $\Theta_{PS} = \Theta_P \cup \Theta_S$, *and the* theory of primary and secondary events with qualification preferences *is the combination of all three theories,* $\Theta_{PQS} = \Theta_P \cup \Theta_Q \cup \Theta_S$.

4.1 Ramifications

As suggested, secondary events provide a means for representing ramifications. This will be illustrated by a series of examples, beginning with a solution to the problem posed by Example 28.

Example 30. Let Θ_{30} consist of the theory of primary and secondary events, Θ_{PS}, and $\Theta_{28} \cup \{42\}$; where:

$$Occ(Move(x,l,l'))(t) \wedge Holds(x,y)(t) \rightarrow \\ Inv(Move(x,l,l'), Move(y,l,l'))(t) \qquad (42)$$

Thus (42) is an invocation axiom which states that in a context in which x is holding y, a move-x event invokes a move-y event.

Then $\Theta_{30} \approx_P At(B,L2)(2)$ as required.

As before $Succ(Move(O,L1,L2))(2)$ is true in all P-preferred models of Θ_{30}. From (32), (33) and the invocation axiom (42), it follows that $Inv(Move(O,L1,L2), Move(B,L1,L2))(1)$ is also true. So it follows by (40) that $Occ(Move(B,L1,L2))(1)$ is true. It follows from (32) and (35) that the precondition of this event, $At(B,L1)(1)$, is true. And it follows from the definition of P-preferred model that $?Qual(Move(B,L1,L2))(1)$ is true. So it follows by the change axiom that $At(B,L2)(2)$. $\square$

Note that, in this example, the introduction of the secondary move event has the effect that the inertia axiom and domain axiom (35) are no longer

[20] Axiom (41) cannot be replaced by invocation axioms of the form: $Succ(e)(t) \wedge \Phi(e,e')(t) \rightarrow Inv(e,e')(t)$, as these would not be interpreted as intended. On the intended use, assuming $?Qual(e)(t)$ results, with (40), in $Occ(e')(t)$. But in P-preferred models events are minimized before qualifications, so in particular $?Occ(e')(t)$ would be preferred to $?Qual(e)(t)$. This is as it should be. The assumption that e succeeds is an assumption about the future and so should not be used to change the present by adding the occurrence of an additional event to it.

used to reason about change. Axiom (35) is now properly used as a domain constraint; for example it requires that, as Ollie is initially holding the block, the block must be at the same location as he is. Moreover the the inertia axiom is now only used to represent inertia; for example, that Ollie is still holding the block after he moves.

The next example illustrates the invocation of a chain of events.

Example 31. Example 30 can be elaborated by stipulating that Ollie is initially holding a stack of blocks. More accurately, suppose that Ollie is holding B and block $B1$ is on top of B and block $B2$ is on top of $B1$. The result of Ollie moving to $L2$ should now be that the stack of blocks goes with him.

Let $\Theta_{31} = \Theta_{30} \cup \{(43), \ldots, (46)\}$ where:

$$On(B2, B1)(1) \wedge On(B1, B)(1) \tag{43}$$

$$On(x, y)(t) \wedge At(y, l)(t) \rightarrow At(x, l)(t) \tag{44}$$

$$Occ(Move(x, l, l')(t) \wedge On(y, x)(t) \rightarrow \\ Inv(Move(x, l, l'), Move(y, l, l'))(t) \tag{45}$$

$$UNA[O, L1, L2, B, B1, B2] \tag{46}$$

Thus domain axiom (44) constrains the relations On and At, and and invocation axiom (45) states that in a context in which y is on x, a move-x invokes a move-y event.

Then, as required, $\Theta_{31} \approx_P At(B1, L2)(2) \wedge At(B2, L2)(2)$.

Proof. Once again, prioritization ensures that potential change is preferred to potential inertia. Thus, as before, B is successfully moved to $L2$, while this time moving B successfully invokes a move-$B1$ event (45), which in turn successfully invokes a move-$B2$ event, (45). $\square$

The following example illustrates the effect of failure on invoked events.

Example 32. Suppose that Example 31 is combined with Example 20. Suppose that is that Stan is initially at $L3$ and that Stan and Ollie attempt to move to $L2$ simultaneously. The result should be chaos; i.e. it should not be possible to say where Stan, Ollie and the blocks end up. Let Θ_{32} be $\Theta_{31} \cup \{(19), (47), \ldots, (50)\}$ where:

$$At(S, L3)(1) \tag{47}$$

$$Occ(Move(S, L3, L2))(1) \tag{48}$$

$$At(O, l)(t) \rightarrow \neg At(S, l)(t) \tag{49}$$

$$UNA[S, O, L1, L2, L3, B, B1, B2] \tag{50}$$

Then there are P-preferred models of Θ_{32} in which Ollie succeeds in moving to $L2$ carrying the blocks with him, while Stan's attempt to move to $L2$ fails, because of axiom (49), and he remains at $L3$.

There are also P-preferred models of Θ_{32} in which Stan succeeds in moving to $L2$ while Ollie's attempt fails, again because of axiom (49). Consequently all of the invoked move-block events fail, because of axiom (41), and he remains holding the blocks at $L1$.

The example can, of course, be further elaborated by adding qualification preferences; as in Example 26. Thus if

$$\Theta'_{32} = \Theta_{32} \cup \Theta_Q \cup \{Qual(Move(O,l,l'), Move(S,l'',l'))(t)\}$$

then $\Theta'_{32} \approx_{PQ} At(O,L2)(2) \wedge At(B_i,L2)(2)$, where B_i is any of the blocks. □

The ramifications considered so far are local, in that they depend on a single event occurring at some stage in an invocation chain, and are thus easily represented as the postconditions of secondary events. We may also want to represent ramifications which are global, in that they depend on the combined result of an indefinite number of events. In order to see the problem, consider the following example:

Example 33. Suppose that the the initial configuration is as in Example 32, but that this time instead of Stan moving a definition of a clear predicate is added. The expected outcome is that Ollie succeeds in moving to $L2$ with all of the blocks, and consequently that $L1$ is clear as a result. Let $\Theta_{33} = \Theta_{31} \cup \{(47),(50),(51),(52),(53)\}$; where:[21]

$$Clear(l)(t) \equiv \neg\exists x At(x,l)(t) \tag{51}$$
$$Object(x) \rightarrow x = S \vee x = O \vee x = B \vee x = B1 \vee x = B2 \tag{52}$$
$$At(x,l)(t) \rightarrow Object(x) \tag{53}$$

Then there are P-preferred models of Θ_{33} in which events unfold as expected. However there are also unintended P-preferred models of Θ_{33} in which $L1$ is not clear at time 2 because Stan mysteriously moves to $L1$. Once again, the unintended models arise because of the interaction between the inertia axiom and a domain axiom, in this case the definition of clear.

Proof. In all P-preferred models of Θ_{33}, prioritization once again ensures that Ollie gets to $L2$ safely with all of the blocks.

[21] Redundant temporal indices in these axioms have been suppressed.

In the intended P-preferred models of Θ_{33}, $?\mathit{Aff}(At(S,L3))(2)$ is true. So it follows by inertia that Stan remains at $L2$. Consequently it follows by (34), (52), and (53) that $\forall x \neg At(x,L1)(2)$. So it follows from (51) that $Clear(L1)(2)$.

In the unintended P-preferred models of Θ_{33}, $?\mathit{Aff}(\neg Clear(L1))(1)$ is true, so it follows by inertia that $\neg Clear(L1)(2)$. Axiom (51) thus requires that $\exists x At(x,L1)(2)$ is true. And axiom (53) requires that x is one of the objects listed in axiom (52). Prioritization has ensured that Ollie and the blocks are safely at $L2$ at time 2. So $\mathit{Aff}(At(S,L3))(1)$ and $At(S,L1)(2)$ are forced. □

The problem raised by this example can be resolved by treating 'clear' as a verb, and thus as an additional secondary event.

Example 34. Let $\Theta_{34} = \Theta_{33} \cup \{(54),(55),(56)\}$ where:

$$Pre(Clear(l))(t) \equiv \neg Clear(l)(t) \tag{54}$$

$$Post(Clear(l))(t) \equiv Clear(l)(t) \tag{55}$$

$$\forall x(At(x,l)(t) \rightarrow \exists l' Occ(Move(x,l,l'))(t)) \rightarrow Inv(Occ(Move(x,l,l'),Clear(l))(t) \tag{56}$$

Thus invocation axiom (56) states that in a context in which a move event occurs for every object at l, each of the move events initiates a clear-l event. It follows, from (41), that the clear event can only succeed if all of the move events succeed; although the move events may succeed and the clear event fail.

Then, as required, $\Theta_{34} \approx_P Clear(L1)(2)$. As in Example 33, Ollie succeeds in moving to $L2$ with all of the blocks. As $\forall x(At(x,L1)(1) \rightarrow Occ(Move(x,L1,L2))(1))$ is true in any P-preferred model of Θ_{34}, it follows from (56) and (40) that $Occ(Clear(L1))(1)$ is true. The precondition of this event is true, and in any P-preferred model of Θ_{34}, $?Qual(Clear(L1))(1)$ is preferred to $?\mathit{Aff}(\neg Clear(L))(1)$, so the conclusion follows by the change axiom.

If the example is extended so that Stan simultaneously attempts to move from $L3$ to $L1$, a conflict arises. This can be resolved by adding the qualification preference $Qual(Move(x,l,l'),Clear(l'))(t)$. □

The missing causal element which accounts for our understanding of Example 33 is an event, and, once again, the choice between events seems clear; intuitively we don't think that Stan moves to $L1$ as this would involve an additional move event, however it seems perfectly natural to say that Ollie clears $L1$ by moving to $L2$ and carrying all of the blocks with him.

4.2 Non-determinism

Non-deterministic events present another variation on the theme of variant effects. It seems that the effects of non-deterministic events can simply be represented as exclusive disjunctions of postconditions, with exclusive disjunction defined as:

$$\phi \,|\, \psi \quad \stackrel{def}{=} \quad (\phi \vee \psi) \wedge \neg(\phi \wedge \psi)$$

However the next example shows that this is insufficient.

Example 35. Suppose that a coin is showing heads initially and that the coin is tossed. Assuming that it is a fair coin, the result should be uncertain; the coin may show heads or it may show tails. Let $\Theta_{35} = \Theta_P \cup \{(57), (58), (59)\}$ where:

$$Heads(1) \wedge Occ(Toss)(1) \tag{57}$$

$$Pre(Toss)(t) \equiv Post(Toss)(t) \equiv Heads(t) \,|\, Tails(t) \tag{58}$$

$$Heads(t) \equiv \neg Tails(t) \tag{59}$$

Then, contrary to intention, $\Theta_{35} \mathrel{|\!\approx_P} Heads(2)$. The toss event succeeds in all P-preferred models of Θ_{35}, and so $Heads(2) \,|\, Tails(2)$ is true. However, $Heads(1)$ and $?\mathit{Aff}(Heads)(1)$ are also true in these models. So the inertia axiom gives $Heads(2)$. □

If, as in this example, one of the possible outcomes of a non-deterministic event maintains the status quo then inertia will favour that effect and so determine the outcome of the event. Once again there seems to be a missing causal element which results in our intuitive understanding of the example but which is missing from the formalization of it. It seems that this time there is not one missing event but *two*. That tossing the coin invokes two conflicting deterministic events; one of which results in the coin showing heads, while the other results in the coin showing tails.

Example 36. Let $\Theta_{36} = \Theta_{35} \cup \Theta_S \cup \{(60), \ldots, (63)\}$ where:

$$Pre(TossH)(t) \equiv Pre(TossT)(t) \equiv Pre(Toss) \tag{60}$$

$$Post(TossH)(t) \equiv Heads(t) \tag{61}$$

$$Post(TossT)(t) \equiv Tails(t) \tag{62}$$

$$Occ(Toss)(t) \rightarrow \\ Inv(Toss, TossH)(t) \wedge Inv(Toss, TossT)(t) \tag{63}$$

Then there are P-preferred models of Θ_{36} in which the toss-heads event succeeds and $Heads(2)$ is true. But there are also P-preferred models of Θ_{36} in which the toss-tails event succeeds (as $?Qual(TossT)(1)$ is preferred to $?Aff(Heads)(1)$) and so $Tails(2)$ is true. □

A similar problem arises in the Russian Shooting Problem [39]. A revolver is loaded with a single bullet and the the cylinder is spun. The result should be uncertain as the cylinder could come to rest with the bullet in any one of six possible positions. However, once again, inertia will favour the outcome in which the bullet remains in its original position. A faithful representation of the outcome of the spin event can be obtained by having it invoke six conflicting secondary spin events, with the result that the bullet is equally likely to come to rest in any of the six possible positions. More generally, a non-deterministic event can be seen as invoking a set of conflicting deterministic events, each resulting in one of its possible outcomes. Thus if e is a non-deterministic event with n distinct outcomes and for each i such that $1 \leq i \leq n$ event e_i results in outcome o_i, then e should invoke the events $e_1, \ldots, e_n$:

$$Occ(e)(t) \rightarrow Inv(e, \{e_1, \ldots, e_n\})(t);$$

where:

$$Inv(e, \{e_1, \ldots, e_n\})(t) \stackrel{def}{=} Inv(e, e_1)(t) \wedge \ldots \wedge Inv(e, e_n)(t).$$

The next example illustrates both non-determinism and ramification.

Example 37. (The Einstein-Podolsky-Rosen paradox) "Suppose that we have a physical system consisting of two sub-systems A and B. For example, take A and B to be two different particles. Suppose that two (orthogonal) alternatives for the state A are $|\alpha\rangle$ and $|\rho\rangle$, whereas B's state might be $|\beta\rangle$ or $|\sigma\rangle$. ... [Then] the general combined state would not simply be a product ('and') of a state A with a state B, but a superposition ('plus') of such products. (We say that A and B are then *correlated.*) Let us take the state of the system to be

$$|\alpha\rangle|\beta\rangle \; + \; |\rho\rangle|\sigma\rangle.$$

Now perform a yes/no measurement on A that distinguishes $|\alpha\rangle$ (**YES**) from $|\rho\rangle$ (**NO**). What happens to B? If the measurement yields **YES**, then the resulting state must be

$$|\alpha\rangle|\beta\rangle,$$

while if it yields **NO**, then it is

$$|\rho\rangle|\sigma\rangle.$$

Thus our measurement of A causes the state of B to jump: to $|\beta\rangle$, in the event of a **YES** answer, and to $|\sigma\rangle$ in the event of a **NO** answer! The particle B need not be localized anywhere near A; they could be light-years apart. Yet B jumps simultaneously with the measurement of A!" [32, p. 281].

The behaviour of this fragment of the quantum world can be represented in the causal theory $\Theta_{37} = \Theta_{PS} \cup \{(64), \ldots, (74)\}$, where:

$$\mathit{State}(A, \alpha|\rho)(t) \wedge \mathit{State}(B, \beta|\sigma)(t) \wedge \mathit{Correlates}(A, B)(t) \tag{64}$$
$$\mathit{Occ}(\mathit{Meas}(A, \alpha|\rho))(1) \tag{65}$$
$$\mathit{State}(s, x|y)(t) \equiv \mathit{State}(s, x)(t) \mid \mathit{State}(s, y)(t) \tag{66}$$
$$\mathit{Correlates}(s, s')(t) \rightarrow \mathit{Correlates}(s', s)(t) \tag{67}$$
$$\mathit{State}(ss', xy)(t) \equiv \mathit{State}(s, x)(t) \wedge \mathit{State}(s', y)(t) \tag{68}$$
$$\begin{aligned}&(\mathit{State}(ss', xy)(t) \mid \mathit{State}(ss', zu)(t)) \equiv \\ &\quad \mathit{Correlates}(s, s')(t) \wedge \mathit{State}(s, x|z)(t) \wedge \mathit{State}(s', y|u)(t)\end{aligned} \tag{69}$$
$$\mathit{Pre}(\mathit{Meas}(s, x|y))(t) \equiv \mathit{Post}(\mathit{Meas}(s, x|y))(t) \equiv \mathit{State}(s, x|y)(t) \tag{70}$$
$$\mathit{Pre}(\mathit{Meas}(s, x))(t) \equiv \exists y \mathit{State}(s, x|y)(t) \tag{71}$$
$$\mathit{Post}(\mathit{Meas}(s, x))(t) \equiv \mathit{State}(s, x)(t) \tag{72}$$
$$\begin{aligned}&\mathit{Occ}(\mathit{Meas}(s, x|y))(t) \rightarrow \\ &\quad \mathit{Inv}(\mathit{Meas}(s, x|y), \{\mathit{Meas}(s, x))(t), \mathit{Meas}(s, y)\})(t)\end{aligned} \tag{73}$$
$$\begin{aligned}&\mathit{Occ}(\mathit{Meas}(s, x|y))(t) \wedge (\mathit{State}(ss', xy)(t) \mid \mathit{State}(ss', zu)(t)) \rightarrow \\ &\quad \mathit{Inv}(\mathit{Meas}(s, x), \mathit{Meas}(s', y))(t) \wedge \mathit{Inv}(\mathit{Meas}(s, z), \mathit{Meas}(s', u))(t)\end{aligned} \tag{74}$$

As required,

$$\Theta_{37} \cup \{\mathit{Succ}(\mathit{Meas}(A, \alpha))(1)\} \approx_P \mathit{State}(A, \alpha)(2) \wedge \mathit{State}(B, \beta)(2),$$
$$\Theta_{37} \cup \{\mathit{Succ}(\mathit{Meas}(A, \rho))(1)\} \approx_P \mathit{State}(A, \rho)(2) \wedge \mathit{State}(B, \sigma)(2).$$

If measuring A results in its state being α then the state of B simultaneously jumps to (is measured as) β, while if measuring A results in ρ then the state of B simultaneously jumps to σ.[22]

[22]Note that the partiality of $\mathcal{TC}$ was not used in representing the initial state of A and B. Thus the initial state of A is represented as $\mathit{State}(A, \alpha|\rho)(1)$, rather than say $?\mathit{State}(A, x)(1)$ together with the constraint $!\mathit{State}(A, x)(t) \rightarrow \mathit{State}(A, \alpha|\rho)(t)$, etc. This is appropriate as "*Both* alternatives must be present in the state, equally weighted as part

Proof. At time 1 the occurrence of the event $Meas(A, \alpha|\rho)$ invokes the events $Meas(A, \alpha)$ and $Meas(A, \rho)$; by (73). Moreover, it follows from (64) and (69) that $State(AB, \alpha\beta)(1) \mid State(AB, \rho\sigma)(1)$. So it follows from (74) that $Meas(A, \alpha)$ invokes $Meas(B, \beta)$ and that $Meas(A, \rho)$ invokes $Meas(B, \sigma)$. Clearly $Meas(A, \alpha|\rho)$ succeeds. Clearly also exactly one of $Meas(A, \alpha)$ and $Meas(A, \rho)$ succeeds; by (64), (66) and (72). Suppose that it is $Meas(A, \alpha)$. Then, by (72), $State(A, \alpha)(2)$. Moreover, as $Meas(A, \rho)$ fails, the event it invokes, $Meas(B, \sigma)$, also fails. So the event invoked by $Meas(A, \alpha)$, $Meas(B, \beta)$, succeeds giving $State(B, \beta)(2)$. Alternatively, if $Meas(A, \rho)$ succeeds then, by analogous reasoning, so does $Meas(B, \sigma)$, with the result that $State(A, \rho)(2)$ and $State(B, \sigma)(2)$. □

5 Change and inertia

The theory of primary and secondary events is realistic insofar as events are causal in character (in that they normally succeed, but may fail), they may be structured (they may invoke other events) and their effects may be variable (their effects may be context-dependent or non-deterministic). This view of events makes it possible to consider a general threefold strategy for the representation of causal reasoning which divides causal theories into three groups of axioms and assigns a clear function to the axioms of each group.

1. The change axioms include the theories of primary events, qualification preferences, and secondary events (Θ_P, Θ_Q, Θ_S), and the definitions of events (axioms which define their preconditions, postconditions and their occurrence). The function of these axioms is to represent events and change. Indeed, all change should result from their use. The

of a quantum linear superposition" [32, p. 292]. Thus there are P-preferred models of Θ_{37} in which $State(A, \alpha)(1)$ is true and P-preferred models of Θ_{37} in which $State(A, \rho)(1)$ is true. Now suppose that measuring the state of A at time 1 results in $State(A, \alpha)(2)$. This has the effect that in those P-preferred models in which $State(A, \rho)(1)$ is true the state of A changes from ρ to α. The effect of measuring A is thus to collapse its state vector to α.

Thus it seems that the present representation of quantum states and events is fairly close to the intended interpretation of quantum theory, and may be useful in explaining it. For example, it seems that the paradox of Schrödinger's cat [32, pp. 290-296] can be resolved by distinguishing between the parallel possibilities represented in the P-preferred models of the appropriate causal theory (the cat being alive in some models and dead in the others), the limited epistemic context of common scientific knowledge (as represented by the set of all P-preferred models of the theory), and the extension of this context which results when opening the door collapses the cat's state vector.

interpretation of these axioms is pragmatic, their proper use depends on appropriate diachronic change assumptions.

2. The inertia axiom forms a group on its own. Its function is to represent inertia and its use should be confined to this purpose; thus it should not be used to infer change. The interpretation of this axiom is pragmatic, its proper use depends on appropriate diachronic inertia assumptions.

3. The third group consists of the remaining axioms, the domain axioms. The function of these axioms is to define the domain and to maintain its integrity. They define the domain by stating which objects it contains and by stating the invariant properties of its relations. They maintain the integrity of the domain by constraining change and inertia assumptions in accordance with these properties. Thus they may prohibit change assumptions (by forcing qualifications) when conflicting events occur, or they may prohibit inertia assumptions (by forcing affectations) when events succeed. Consequently their interpretation should be semantic.

The methodological undesirability of using the inertia axiom in order to reason about change has already been noted and illustrated in examples 28 and 33. And this has been eliminated by the addition of secondary events which achieve the intended secondary effects (as in examples 30 and 34).

This restriction of the inertia axiom seems entirely appropriate. But the suggested separation of the functions of the change and domain axioms may seem less so, as it means that all changes have to be specified in the postconditions of events. However, given the present view of events, it seems that this can be done; and indeed it has been done in all of the successful examples considered thus far. It is also methodologically desirable. Domain axioms should, as suggested, be used to define the domain and to maintain its integrity. For example (34) prevents an object being at two different locations simultaneously (it can thus be thought of as imposing a metaphysical constraint) and (51) defines what it means for a location to be clear (and can thus be thought of as a meaning postulate). Rules of this sort are declarative and so should be used to describe the world, not to change it.

However if the strategy is to work, a proper balance needs to be found between reasoning about change and reasoning about inertia. Thus far the emphasis has been on constraining the inertia axiom, first by refining the pragmatics so that change is preferred to inertia (Example 14), and then by adding secondary events which overcome its effects (as in the examples referred to above and examples 35, 36) and 37. As a result change always dominates inertia.

But this dominance is not always be desirable. Consider again the problem posed by the global *Clear* ramification in Example 33. Unlike the local ramifications of earlier examples, which can be represented as the secondary effects of individual events, the *Clear* ramification depends on the outcome of an indefinite number of events. This is reflected in the solution given in Example 34, where a complex invocation rule is used to ensure that a secondary clear-l event is only invoked if all of the objects at l are moved, and which can only succeed if all of the move events do.

It would be preferable if global ramifications such as this could be treated locally; if, in this case, each move-from-l event could individually initiate a secondary clear-l event and the consequences could be left to take care of themselves.

However if the antecedent of the invocation axiom is weakened to allow this then the clear-l event becomes too strong and unwanted changes result.

Example 38. Suppose that Ollie is at location $L1$ initially and that he is holding block $B1$. Suppose also that block $B2$ is at $L1$. Then if Ollie moves to location $L2$ the result should be that Ollie arrives safely at $L2$, taking $B1$ with him, and that $B2$ remains at $L1$.

But if $\Theta_{38} = \Theta_{PS} \cup \{(75), \ldots, (86)\}$ where:[23]

$$At(O, L1)(1) \land Holds(O, B1)(1) \land At(B2, L1)(1) \tag{75}$$

$$At(x, l)(t) \land l \neq l' \rightarrow \neg At(x, l')(t) \tag{76}$$

$$At(x, l)(t) \land Holds(x, y)(t) \rightarrow At(y, l)(t) \tag{77}$$

$$Clear(l)(t) \equiv \neg \exists x At(x, l)(t) \tag{78}$$

$$UNA[O, B1, B2, L1, L2] \tag{79}$$

$$Occ(Move(O, L1, L2))(1) \tag{80}$$

$$Pre(Move(x, l, l'))(t) \equiv At(x, l)(t) \tag{81}$$

$$Post(Move(x, l, l'))(t) \equiv At(x, l')(t) \land \neg At(x, l)(t) \tag{82}$$

$$Pre(Clear(l))(t) \equiv \neg Clear(l)(t) \tag{83}$$

$$Post(Clear(l))(t) \equiv Clear(l)(t) \tag{84}$$

[23] Axioms given earlier are repeated here in order to improve readability. The axioms have also been arranged in groups.

$$Occ(Move(x,l,l'))(t) \wedge Holds(x,y)(t) \rightarrow \\ Inv(Move(x,l,l'), Move(y,l,l'))(t) \quad (85)$$

$$Occ(Move(x,l,l'))(t) \wedge At(x,l)(t) \rightarrow \\ Inv(Move(x,l,l'), Clear(l))(t) \quad (86)$$

Then $\Theta_{38} \approx_P Clear(L1)(2)$.

Proof. As before, Ollie's moving invokes $B1$'s moving. Both of these events now invoke a clear-$L1$ event, by the local invocation rule (86). As change is preferred to inertia, all three events succeed. So, in particular, $Clear(L1)(2)$ is true in all P-preferred models of Θ_{38}. It then follows from the definition of clear, (78), that $\neg At(B2, L1)(1)$ and consequently, by the inertia axiom, that $Aff(At(B2, L1))(1)$. As a result of the clear-$L1$ event, $B2$ has mysteriously become locationless. □

The situation is aggravated if the domain is represented more accurately.

Example 39. In Example 38, the effects of the clear-$L1$ event violate the integrity of the domain, particularly the intuition that "everything must be somewhere", that each object must always be at some location or other. This constraint is maintained by the move event, and can be imposed generally by adding the domain axioms:[24]

$$Object(x) \rightarrow x = O \vee x = B1 \vee x = B2 \quad (87)$$

$$Location(x) \rightarrow x = L1 \vee x = L2 \quad (88)$$

$$Object(x) \rightarrow \exists l(Location(l) \wedge At(x,l)(t)) \quad (89)$$

Then $\Theta_{39} = \Theta_{38} \cup \{(87), (88), (89)\}$ still predicts $Clear(L1)(2)$. However Θ_{39} also predicts $At(B2, L2)(2)$. The more accurate representation of the domain has the further unintended consequence (by axioms (87)-(89)) that clearing $L1$ now results in $B2$ mysteriously moving to $L2$. □

Clearly a better balance between change and inertia, which tempers effects of events such as $Clear$, is desirable. Fortunately this can be achieved, given an appropriate definition of causation.

6 Causation

Hume famously raised the problem of defining causation [17, Book I, Part III]. The assertion that event e caused event e' seems to claim more than

[24] Redundant temporal indices in these axioms have been suppressed.

that events of type e are regularly succeeded by events of type e'. But what more is being claimed? Mackie observes that it cannot be that e is both a necessary and sufficient preceding condition for e':

> Suppose that a fire has broken out in a certain house, but has been extinguished before the house has been completely destroyed. Experts investigate the cause of the fire and they conclude that it was caused by an electrical short-circuit at a certain place. What exactly is the force of their claim that this short-circuit caused this fire? Clearly the experts are not saying that the short-circuit was a necessary condition for this house's catching fire at this time. They know perfectly well that a short-circuit somewhere else, or the overturning of a lighted oil stove, or any one of a number of other things might, if it had occurred, have set the house on fire. Equally, they are not saying that the short-circuit was a sufficient condition for this house's catching fire; for if the short-circuit had occurred, but there had been no inflammable material near by, the fire would not have broken out, and even given both the short-circuit and the inflammable material, the fire would not have occurred if, say, there had been an efficient automatic sprinkler at just the right spot. Far from being a condition both necessary and sufficient for the fire, the short-circuit was, and is known to the experts to have been, neither necessary nor sufficient for it. In what sense, then, is it said to have caused the fire? [26, pp. 15-16]

Mackie's answer is that the short-circuit is said to have caused the fire because the short-circuit was an INUS condition for the house's catching fire:

> At least part of the answer [to the question of what caused the fire] is that there is a set of conditions (of which some are positive and some are negative), including the presence of inflammable material, the absence of a suitably placed sprinkler, and no doubt quite a number of others, which combined with the short-circuit constituted a complex condition that was sufficient for the house's catching fire—sufficient but not necessary, for the fire could have started in other ways. Also, of *this* complex condition, the short-circuit was an indispensable part: the other parts of this condition, conjoined with one another in the absence of the short-circuit, would not have produced the fire. The short-circuit which is said to have caused the fire is thus

> an indispensable part of a complex sufficient (but not necessary) condition of the fire. In this case, then, the so-called cause is, and is known to be, an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result. The experts are saying, in effect, that the short-circuit is a condition of this sort, that it occurred, that the other conditions which conjoined with it form a sufficient condition were also present, and that no other sufficient condition of the house's catching fire was present on this occasion. I suggest that when we speak of the cause of some particular event, it is often a condition of this sort that we have in mind. In view of the importance of conditions of this sort in our knowledge of and talk about causation, it will be convenient to have a short name for them: let us call such a condition (from the initial letters of the words italicized above), an INUS condition. [p. 16]

A formal definition of causation along these lines is now developed. The relation of direct causation, *Cause*, holds between sentences of the form $Occ(e)(t)$ and sentences of $\mathcal{TC}$. The definition is then extended to indirect causation by taking the transitive closure, *Causes*, of *Cause*. The intended reading of a sentence of the form $Cause(Occ(e)(t), \phi)$ is thus that the occurrence of event e at time t is the (direct) cause of ϕ, that the occurrence of e at t is is an INUS condition for ϕ.

The definition assumes the setting of a finite causal theory $\Theta = \Theta_{PS} \cup \Theta_B$, where Θ_{PS} is the familiar theory of primary and secondary events and Θ_B is the background causal theory, consisting of change and domain axioms.

The definition begins with sufficient causation, with the definition of the notion of a cause as part of an unnecessary but sufficient condition for the effect. It seems plausible to say that the occurrence of an event e at time t is sufficient to cause ϕ if e succeeds and $Post(e)(t+1) \equiv \phi$. However this has the unfortunate consequence that if e succeeds at t then e is a sufficient cause of any sentence ϕ which is true at $t+1$. Clearly what is intended is not just that the sentence $Post(e)(t+1) \equiv \phi$ is true, but that it is always true given the definition of the postconditions of e; that it is a semantic consequence of the postconditions of e in $\mathcal{TC}$.

In order to express the notion of semantic consequence in the object language. the necessity operator, $\Box$, is added to $\mathcal{TC}$. Informally, for $\mathcal{TC}$-sentence ϕ, $\Box\phi$ states that ϕ is necessarily true, that ϕ is a tautology of $\mathcal{TC}$. As the use of this operator is limited, the technical changes will be kept to a minimum. The formal syntax of $\mathcal{TC}^{\Box}$ is a straightforward extension of the syntax of $\mathcal{TC}$: for any $\phi \in \mathcal{TC}$, $\{\phi, \Box\phi\} \subseteq \mathcal{TC}^{\Box}$. The semantics of $\mathcal{TC}$ is also extended in a straightforward way. Models of $\mathcal{TC}^{\Box}$ are simply models of $\mathcal{TC}$ and Definition 7 is extended to $\mathcal{TC}^{\Box}$ by adding the clauses:

$$\begin{array}{lll} M, g \models \Box\psi & \textit{iff} & M', g' \models \psi \textit{ for all } M' \textit{ and } g' \\ M, g \mathrel{=\!\!|} \Box\psi & \textit{iff} & M', g' \mathrel{=\!\!|} \psi \textit{ for some } M' \textit{ and } g' \end{array}$$

This definition is unusual in that the satisfaction (violation) of a formula of the form $\Box\psi$ by assignment g in model M may depend on the satisfaction (violation) of ψ by other assignments in other models. Nevertheless the definition is adequate for the purpose, and it has the advantage that it obviates a full-blown modal semantics. Finally, the pragmatics of causal theories remains unchanged; except of course for that it now applies to $\mathcal{TC}^{\Box}$ and so assumes the above extensions.

Now, as Θ is assumed to be finite, the postcondition definitions in Θ_B can be expressed as the $\mathcal{TC}$-sentence:

$$\mathit{Post}(\Theta_B) = \bigwedge\{\forall x_1, \ldots, x_n, t(\mathit{Post}(f(x_1, \ldots, x_n))(t) \equiv \phi) \in \Theta_B : f \in F_E\}$$

So the first part of the definition of sufficient causation can now be expressed as follows:[25]

$$\mathit{Succ}(e)(t) \wedge \Box(\mathit{Post}(\Theta_B) \wedge \mathit{Post}(e)(t+1) \rightarrow \phi) \rightarrow \mathit{SCause}(\mathit{Occ}(e)(t), \phi)$$

Thus the occurrence of event e at time t is sufficient to cause ϕ if e succeeds at t and ϕ is a semantic consequence of the postconditions of e at $t+1$. The occurrence of e. When this is the case, the occurrence of e at t can be regarded as being part of an unnecessary but sufficient condition for ϕ. The whole of the sufficient condition is the conjunction $\mathit{Pre}(e)(t) \wedge \mathit{Occ}(e)(t) \wedge \bullet \mathit{Qual}(e)(t)$. So the positive conditions required for the success of the event are its preconditions and the negative conditions are represented by the absence of qualifications. It is thus possible to avoid giving a fixed list of negative conditions, which will always be inadequate as another negative condition (another qualification) can always arise, while at the same time vagaries such as "and no doubt quite a number of others" are eliminated.

The second part of the definition of sufficient causation is more straightforward. The occurrence of an event e at time t is a sufficient cause of any event e' which it successfully invokes at t:

$$\mathit{Inv}(e, e')(t) \wedge \mathit{Succ}(e')(t) \rightarrow \mathit{SCause}(\mathit{Occ}(e)(t), \mathit{Occ}(e')(t))$$

[25] From a technical point of view, the following formula should be viewed as a $\mathcal{TC}^{\Box}$ sentence schema in which $\mathit{Post}(\Theta_B)$ is a variable which ranges over $\mathcal{TC}$-sentences. The schema becomes a $\mathcal{TC}^{\Box}$-sentence when the background theory Θ_B is supplied.

The requirement that e' succeeds restricts the definition of sufficient causation to events which are causally significant; and, of course, e' can only succeed if e does (41).

Combining the two parts gives the definition of sufficient causation:

$$\begin{aligned}
&SCause(Occc(e)(t), \phi) \equiv \\
&\quad (Succ(e)(t) \wedge \Box(Post(\Theta_B) \wedge Post(e)(t+1) \rightarrow \phi)) \vee \\
&\quad (Inv(e, \phi)(t) \wedge Succ(\phi)(t))
\end{aligned} \tag{90}$$

Turning now to the definition of necessary causation, to the definition of the notion of a cause being an insufficient but necessary condition. The occurrence of e at t is necessary for ϕ if no other event e' which occurs at t and which is not invoked by e is sufficient for ϕ:

$$\begin{aligned}
&NCause(Occ(e)(t), \phi) \equiv \\
&\quad \bullet\exists e'(e \neq e' \wedge \bullet Inv(e, e')(t) \wedge SCause(Occ(e')(t), \phi))
\end{aligned} \tag{91}$$

Thus if e invokes e' at t and both are sufficient causes of ϕ, then e may still be deemed necessary for ϕ as e' depends on e. This proviso is illustrated in Example 43 below.

Combining these two definitions thus gives an INUS definition of causation:

$$\begin{aligned}
&Cause(Occ(e)(t), \phi) \equiv \\
&\quad SCause(Occ(e)(t), \phi) \wedge NCause(Occ(e)(t), \phi)
\end{aligned} \tag{92}$$

The definition can now be extended to indirect causation:

$$\begin{aligned}
&Causes(Occ(e)(t), \phi) \equiv \\
&\quad Cause(Occ(e)(t), \phi) \vee \\
&\quad \exists e', t'(Cause(Occ(e)(t), Occ(e')(t')) \wedge Causes(Occ(e')(t'), \phi))
\end{aligned} \tag{93}$$

The *Causes* relation has the following properties:

$$\neg Causes(\phi, \phi) \tag{94}$$

$$Causes(\phi, \psi) \wedge Causes(\psi, \chi) \rightarrow Causes(\phi, \chi) \tag{95}$$

$$Causes(\phi, \psi) \rightarrow \neg Causes(\psi, \phi) \tag{96}$$

$$Causes(\phi, \psi) \rightarrow \phi \wedge \psi \tag{97}$$

$$Causes(\phi, \psi) \rightarrow \neg Causes(\phi, \bullet\psi) \tag{98}$$

$$Causes(\phi, \psi) \wedge \Box(\psi \rightarrow \chi) \rightarrow Causes(\phi, \chi) \tag{99}$$

Thus *Causes* is irreflexive, (94), as *Inv* is, and transitive, (95). Consequently it is also asymmetric, (96). Property (97) states that causation is actual, that if ϕ causes ψ then both ϕ and ψ are true. This implies the consistency of causation, (98); which holds because events with mutually inconsistent postconditions (including events with inconsistent postconditions) give rise to qualifications and thus fail. Finally, *Causes* is closed under semantic consequence, (99).

The definition of *Causes* makes it possible to talk more clearly about causation when discussing examples.

Example 40. Recall that in the dominoes example, Example 19, the toppling of domino d_1 at time 1 results in the toppling of domino d_{100} at time 100. Clearly it seems appropriate to say that the former event caused the latter event. And, indeed, if $\Theta_{40} = \{(90), \ldots, (93)\}$, then $\Theta_{19} \cup \Theta_{40} \approx_P$ $\mathit{Causes}(\mathit{Occ}(\mathit{Topple}(d_1, l_1))(1), \mathit{Occ}(\mathit{Topple}(d_{100}, l_{100}))(100))$. This is easily verified by noting that for each $n < 100$, the occurrence of the event $\mathit{Topple}(d_n, l_n)$ at time n is both a sufficient and a necessary cause of the event $\mathit{Topple}(d_{n+1}, l_{n+1})$ at $n{+}1$. The occurrence of $\mathit{Topple}(d_n, l_n)$ at time n is sufficient because it succeeds and $\mathit{Topple}(d_{n+1}, l_{n+1})$ is a semantic consequence of its postconditions at $n+1$; $\mathit{Post}(\Theta_B)$ is just the sentence (14), and clearly $\Box((14) \wedge \mathit{Post}(\mathit{Topple}(d_n, l_n))(n{+}1) \to \mathit{Occ}(\mathit{Topple}(d_{n+1}, l_{n+1}))(n{+}1))$. The occurrence of $\mathit{Topple}(d_n, l_n)$ at time n is necessary because it is the only event which occurs at n; as a result of the minimization of events (*Occ*-literals). $\Box$

Example 41. Similarly in the quantum theory example, Example 37, measuring the state of A causes the state of B to jump: to β in the event that A is in state α, and to σ in the event that A is in state ρ. And, indeed,

$$\begin{array}{ll} \Theta_{37} \cup \Theta_{40} \approx_P & \mathit{Causes}(\mathit{Occ}(\mathit{Meas}(A, \alpha|\rho))(1), \mathit{State}(B, \beta)(2)) \quad | \\ & \mathit{Causes}(\mathit{Occ}(\mathit{Meas}(A, \alpha|\rho))(1), \mathit{State}(B, \sigma)(2)) \end{array}$$

Clearly, at time 1, the occurrence of $\mathit{Meas}(A, \alpha|\rho)$ is the direct cause (the *Cause*) of $\mathit{Meas}(A, \alpha)$ or of $\mathit{Meas}(A, \rho)$. Suppose the former. Then, as before, the success of $\mathit{Meas}(A, \alpha)$ results in $\mathit{State}(B, \beta)(2)$, and so clearly $\mathit{Causes}(\mathit{Occ}(\mathit{Meas}(A, \alpha|\rho))(1), \mathit{State}(B, \beta)(2)$; $\mathit{Occ}(\mathit{Meas}(A, \alpha|\rho))(1)$ is the direct cause of $\mathit{Occ}(\mathit{Meas}(A, \alpha))(1)$ which in turn is the direct cause of $\mathit{Occ}(\mathit{Meas}(A, \beta))(1)$, which directly causes $\mathit{State}(B, \beta)(2)$. Moreover, the success of $\mathit{Meas}(A, \alpha)$ results in the failure of $\mathit{Meas}(A, \rho)$; as their postconditions conflict. As the preconditions of $\mathit{Meas}(A, \rho)$ are true (at time 1) it follows that it is false that $\mathit{Meas}(A, \rho)$ succeeds; that is, $\neg \mathit{Succ}(\mathit{Meas}(A, \rho))(1)$.

And so it follows from the definition of *Cause* that it is false that $Meas(A, \alpha|\rho)$ is the direct cause of $Meas(A,\rho)$; that is, $\neg Cause(Occ(Meas(A,\alpha|\rho))(1), Occ(Meas(A,\rho))(1))$. So it follows from the minimization of events that $Meas(A,\alpha|\rho)$ is not an indirect cause of $State(B,\sigma)(2)$. And clearly $Meas(A, \alpha|\rho)$ is not a direct cause of $State(B,\sigma)(2)$. So it is false that $Occ(Meas(A,\alpha|\rho)$ *Causes* $State(B,\sigma)(2)$; $\neg Causes(Occ(Meas(A,\alpha|\rho))(1), State(B,\sigma)(2))$. A similar argument shows that if instead $Meas(A,\alpha|\rho)$ directly causes $Meas(A, \rho)$, then it also causes $State(B,\sigma)(2))$ and does not cause $State(B,\beta)(2)$. □

The definition of causation also makes it possible to state a Law of Change and Inertia:

$$Aff(\ell)(t) \equiv \ell(t) \wedge \exists e Cause(Occ(e)(t), \bullet\ell(t+1)) \tag{100}$$

Thus the law states that Kleene literal ℓ is affected at time t if and only if ℓ is true at t and some event causes ℓ to have a different truth value at $t+1$. Less formally, the law states that nothing changes without a cause, or that every change has a cause.

Definition 42. *The definition of Causes together with the Law of Change and Inertia form the* theory of causation, $\Theta_C = \{(90), \ldots, (93), (100)\}$. *Combinations of theories are denoted by juxtaposition of their subscripts. Thus* $\Theta_{PSC} = \Theta_P \cup \Theta_S \cup \Theta_C$, *etc.*

The Law of Change and Inertia can be used to regulate the use of the change and inertia axioms, and thus ensure the proper balance between change and inertia.

Clearly the law's presence regulates the inertia axiom, as it does not permit it to be used with domain axioms to infer uncaused effects. So, for example, the unintended use of the inertia axiom in Example 33 would be deemed illegal as it results in Stan's uncaused and mysterious move.

The law's presence also regulates the change axiom as it does not allow an event to succeed if this would result in uncaused effects. Thus it seems that secondary events, such as the *Clear* event, can be invoked locally and that the consequences *can* be left to take care of themselves.

Example 43. Suppose that the theory of causation is added to the theory of Example 38. Suppose, that is, that $\Theta_{43} = \Theta_{38} \cup \Theta_C$. Then, as desired, $\Theta_{43} \approx_P At(Ollie, L2)(2) \wedge At(B1, L2)(2) \wedge At(B2, L1)(2)$. Thus Ollie's moving causes $B1$ to move to $L2$, as he is holding $B1$ at the time. But his moving does not cause $B2$ to move, so $B2$ remains at $L1$.

Proof. As before, both of the move events succeed. The success of the move-Ollie event results in two changes:

$$Aff(At(O, L1))(1) \text{ and } Aff(\neg At(O, L2))(1).$$

But clearly Ollie's moving causes these two changes; that is:

$$\begin{aligned}&Cause(Occ(Move(O, L1, L2))(1), \neg At(O, L1)(2)), \text{ and}\\ &Cause(Occ(Move(O, L1, L2))(1), At(O, L2)(2))\end{aligned}$$

are true. So the Law of Change and Inertia is satisfied. Similarly the success of the move-$B1$ event results in two changes: $Aff(At(B1, L1))(1)$ and $Aff(\neg At(B1, L2))(1)$. But, as before, it follows that:

$$\begin{aligned}&Cause(Occ(Move(B1, L1, L2))(1), \neg At(B1, L1)(2)), \text{ and}\\ &Cause(Occ(Move(B1, L1, L2))(1), At(B1, L2)(2)).\end{aligned}$$

So, again, the law is satisfied. However, the law forces the failure of the clear-$L1$ event. This is because its success would result in two changes: $Aff(\neg Clear(L1))(1)$ and $Aff(At(B2, L1))(1)$. But, although the success of the clear event would cause $Clear(L1)(2)$, it would not cause $\bullet At(B2, L1)(2)$. Consequently the law would require some other cause. It follows from the definition of causation that an additional event would be required as the cause. However in P-preferred models events are minimized before qualifications. Consequently the failure of the clear-$L1$ event is preferred to the occurrence of an additional, mysterious, cause.[26] □

Example 44. However if it is also the case that $B2$ is on $B1$ initially:

$$On(B2, B1)(1) \tag{101}$$

$$Occ(Move(x, l, l'))(t) \wedge On(y, x)(t) \rightarrow Inv(Move(x, l, l'), Move(y, l, l'))(t) \tag{102}$$

then Ollie's moving should clear $L1$. And, indeed, $\Theta_{43} \cup \{(101), (102)\} \approx_P Clear(L1)(2)$. Thus, as $B2$ is on $B1$, Ollie's moving now also causes $B2$ to move to $L2$ and and it causese $L1$ to be clear.

Proof. Ollie's moving invokes a move-$B1$ event and now, as $B2$ is on $B1$, the move-$B1$ event invokes a move-$B2$ event. This causes the two changes

[26] The case for the extended theory $\Theta_{39} \cup \Theta_C$ is similar, except that this time the clear-$L1$ event also fails because its success would result in the additional uncaused change $Aff(\neg At(B2, L2))(1)$.

it introduces; $\mathit{Aff}(At(B2, L1))(1)$ and $\mathit{Aff}(\neg At(B2, L2))(1)$. Consequently it satifies the law. Moreover as the move-$B2$ event now causes the change $\mathit{Aff}(At(B2, L1))(1)$, and the success of the clear-$L1$ event would cause the the only remaining change; $\mathit{Aff}(\neg Clear(L1))(1)$. So the clear-$L1$ event now meets the requirements of the law, and so it is allowed to succeed.

All three move events invoke the clear-$L1$ event. However, given the proviso in the definition of necessary causation, (91), it follows that Ollie's moving counts as its cause; $Causes(Occ(Move(O, L1, L2))(1), Clear(L1)(2))$.

□

As can be seen from Example 43, the Law of Change and Inertia prevents an event from succeeding if this would result in uncaused changes. If the success of an event e would result in an uncaused change, the law requires the occurrence of at least one additional event as the cause. But the pragmatics of causal theories prefer the failure of e to the occurrence of an additional event. This can be seen as further justification for the pragmatics; speculations about the future (in this case that e succeeds) should not be allowed to alter the present (by requiring the occurrence of an additional event or events).

Consequently the law has the effect that while *caused* change is preferred to inertia, inertia is preferred to *uncaused* change. This results in a Yin/Yang interplay between the opposite but complementary principles of change and inertia. The principle of change appears to be dominant; as change is preferred to inertia and events can be initiated whenever they might succeed. However if the success of an event would give rise to changes which it did not cause, the event fails, and inertia prevails.

7 Causal counterfactuals

This paper has been concerned with the representation of actual causal reasoning. The next step is to extend the theory to counterfactual causal reasoning in order to be able to represent causal reasoning about non-actual worlds. A brief account of how this will be done is given in this section.

The Temporal Calculus is extended to the Modal Temporal Calculus, $\mathcal{MTC}$, by adding the conditional operator $\Rightarrow$. Thus Definition 3 is modified by replacing all occurrences of $\mathcal{TC}$ with $\mathcal{MTC}$ and adding the clause:

- *If* $\phi, \psi \in \mathcal{MTC}$ *then* $(\phi \Rightarrow \psi) \in \mathcal{MTC}$.

A sentence of the form $\phi \Rightarrow \psi$ states that if ϕ were the case, then ψ would be the case also; that is, that $\phi \wedge \bullet\psi$ is, in some sense, a remoter possibility than $\phi \wedge \psi$.

A model for $\mathcal{MTC}$ is a pair $\langle W, \prec \rangle$, where W is a non-empty set of $\mathcal{TC}$-models and $\prec$ is a strict partial order on W; that is, $\prec$ is a binary relation on W which is irreflexive and transitive. Intuitively W is thought of as a set of partial possible worlds (which will henceforth be referred to simply as 'worlds') and $\prec$ represents the closeness relation between worlds; thus if $w \prec w' \prec w''$ then w' is, in the appropriate sense, closer to w than w'' is. The reflexive closure of the closeness relation is defined in the usual way: $w \preceq w'$ if and only if either $w \prec w'$, or $w \in W$ and $w = w'$.

The central semantic notion is that of the truth (falsity) of a sentence ϕ at a world w in a model M. This is defined by means of the intermediary notions of satisfaction (violation) of a formula by a variable assignment g at a world w in a model M, written $M, w, g \models \phi$ $(M, w, g \mathrel{=\!\!|} \phi)$. Where possible the definition follows that for $\mathcal{TC}$; Definition 7. Thus, for example:

$$
\begin{array}{rll}
M, w, g \models r(u_1, \ldots, u_n)(t) & \textit{iff} & w, g \models r(u_1, \ldots, u_n)(t) \\
M, w, g \mathrel{=\!\!|} r(u_1, \ldots, u_n)(t) & \textit{iff} & w, g \mathrel{=\!\!|} r(u_1, \ldots, u_n)(t) \\
M, w, g \models \phi \wedge \psi & \textit{iff} & M, w, g \models \phi \textit{ and } M, w, g \models \psi \\
M, w, g \mathrel{=\!\!|} \phi \wedge \psi & \textit{iff} & M, w, g \mathrel{=\!\!|} \phi \textit{ or } M, w, g \mathrel{=\!\!|} \psi
\end{array}
$$

The clauses for the new conditional operator are as follows:

$$
\begin{array}{rll}
M, w, g \models \phi \Rightarrow \psi & \textit{iff} & \textit{for any } w' \textit{ such that } w \preceq w' \textit{ and } M, w', g \models \phi \\
& & \quad \textit{there is a } w'' \textit{ such that} \\
& & \qquad w \preceq w'' \preceq w' \textit{ and } M, w'', g \models \phi \wedge \psi \\
M, w, g \mathrel{=\!\!|} \phi \Rightarrow \psi & \textit{iff} & \textit{there is a } w' \textit{ such that } w \preceq w' \textit{ and } M, w', g \models \phi \\
& & \quad \textit{and there is no } w'' \textit{ such that} \\
& & \qquad w \preceq w'' \preceq w' \textit{ and } M, w'', g \models \phi \wedge \psi
\end{array}
$$

Then a sentence ϕ is true at a world w in a model M (written $M, w \models \phi$) if and only if $M, w, g \models \phi$ for all assignments g, and ϕ is false at w (written $M, w \mathrel{=\!\!|} \phi$) if and only if $M, w, g \mathrel{=\!\!|} \phi$ for every assignment g.

So the counterfactual sentence $\phi \Rightarrow \psi$ is true at a world w if every $\preceq$-closest ϕ-world to w is also a ψ-world, and $\phi \Rightarrow \psi$ is false at w if some $\preceq$-closest ϕ-world is not a ψ-world.[27]

[27] The truth and falsity conditions are more general than the informal characterisation of them suggests; as they give the intended result even when there are infinitely descending chains of worlds above w and hence no closest worlds to w. For present purposes it can be assumed that the Limit Assumption [20] holds—that no infinitely descending chains are involved—and consequently that it is safe to talk of the closest ϕ-worlds to w.

The necessity operator of $\mathcal{TC}^{\Box}$ can be defined in $\mathcal{MTC}$ as follows:

$$\Box\phi \stackrel{def}{=} \bullet\phi \Rightarrow \phi$$

Thus $\Box\phi$ is true at world w just in case ϕ is true at every world w' such that $w \prec w'$.

The semantics for the counterfactual conditional are fairly standard; for example the truth condition is analogous to the one proposed by Veltman in a classical possible worlds setting [49]. Indeed, the semantics of counterfactuals are well understood. Such differences as there are tend to concern the properties of the closeness relation; for example it is additionally required to be almost-connected by Lewis [20], and connected by Stalnaker [45].

The real problem of counterfactuals is not semantic but pragmatic: how should particular closeness relations be defined? Fortunately for the counterfactuals of current interest, for *causal counterfactuals*, an answer has already been prepared: the closeness relation, $\prec$, can be identified with the preference relation for causal theories, $\prec_P$.[28] Thus an $\mathcal{MTC}$-model $\langle W, \prec\rangle$ in which every world in W is a Θ_{PS}-world and $\prec$ is $\prec_P$ will be called a *causal model.* Causal models thus provide an pragmatics for causal counterfactuals. They can also be used for prediction. A causal theory Θ predicts ϕ in a causal model M (written $\Theta \approx_P^M \phi$) if and only if ϕ is true at all P-preferred Θ-worlds in M. The result is an interesting combination of counterfactuals and causal theories.

By way of illustration consider the representation of explanatory causal reasoning. This gives rise to two problems which can be called the postponement problem and the generation problem. The postponement problem was raised by Kautz [18] and is commonly believed to be a fatal weakness of approaches based on chronological minimization. As chronological minimization has the effect of postponing change for as long as possible it cannot, it is claimed, be used to represent explanatory causal reasoning. Kautz gives the example of a car which is in a parking lot at time 1, but which is not longer there when its owner returns at time 4. The plausible explanation is that the car was stolen at some time between time 1 and time 3. However attempts to formalize this using chronological minimization result in the steal event being postponed for as long as possible, and so they yield the counter-intuitive explanation that the car must have been stolen at time 3. The generation problem is more general and faces any formal theory which is designed to represent predictive causal reasoning. Prediction is concerned

[28] Or, when appropriate, with the preference relation for causal theories with qualification preferences, $\prec_{PQ}$.

with drawing reasonable conclusions about the future on the basis of the information currently available. Reasoning of this kind depends on making appropriate conjectures about the future, conjectures that given events will succeed or that given facts will persist. However when reasoning predictively it is important not to go beyond what can be reasonably concluded and suppose that additional events occur or that additional facts are true. By contrast, explanation is concerned with explaining why some fact or event is the case, with the conjecturing of additional facts or events which would account for it. Consequently formal theories of predictive reasoning may fail to generate the additional information required in explanations.

However it seems that causal counterfactuals can be used to overcome both of these problems. The definition of P-preferred minimization has the effect that as one proceeds up any chain of worlds above a P-preferred Θ-world, w, one encounters worlds in which additional literals are true. Moreover these additions arise antichronologically; the higher one ascends, the earlier the additions. Thus the worlds above w represent the various ways in which the history of w might differ as a result of changes introduced at progressively earlier moments. So, as causal counterfactuals can be used to refer to these worlds, they can be used to provide the desired explanations.

This use of counterfactuals is illustrated by the following definition:[29]

$$
\begin{aligned}
&\mathit{Expl}(\mathit{Occ}(e)(t), \phi(t')) \equiv \\
&\qquad t < t' \wedge \\
&\qquad (\mathit{Occ}(e)(t) \Rightarrow (\mathit{Cause}(\mathit{Occ}(e)(t), \phi(t+1)) \wedge \\
&\qquad\qquad\qquad \bullet\exists e', t''(t < t'' < t' \wedge \mathit{Cause}(\mathit{Occ}(e')(t''), \bullet\phi(t''+1))))
\end{aligned}
\tag{103}
$$

So the occurrence of event e at time t is an explanation of ϕ at the later time t' if and only if in the closest worlds in which e occurs at t: e causes ϕ at $t+1$, and no subsequent event e' occurring before time t' causes $\bullet\phi$. Note that a weaker notion of explanation based on sufficient causation can be provided by by using *SCause* instead of *Cause*. Assuming that there are a finite number of explanations, they be collected together as a disjunction:

$$
\mathit{Expls}(\phi, \psi(t)) \equiv (\phi \equiv \bigvee\{\mathit{Occ}(e)(t') : \mathit{Expl}(\mathit{Occ}(e)(t'), \psi(t))\}) \tag{104}
$$

These two definitions form the theory of explanation, Θ_E.

---

[29] $\phi(t)$ denotes any formula in which the temporal variable t occurs free.

Example 45. The stolen car problem can be represented by the theory $\Theta_{45} = \Theta_{PSC} \cup \Theta_E \cup \{(105), (106), (107)\}$; where:

$$In(1) \wedge \neg In(4) \tag{105}$$

$$Pre(Steal)(t) \equiv In(t) \tag{106}$$

$$Post(Steal)(t) \equiv \neg In(t) \tag{107}$$

Let M be the causal model in which W is the set of all Θ_{45}-worlds. Then, as desired, $\Theta_{45} \approx_P^M$ $Expls(Occ(Steal)(1) \vee Occ(Steal)(2) \vee Occ(Steal)(3), \neg In(4))$.

Proof. Let w be a P-preferred Θ_{45}-world in M. Then the closest worlds to w at which $Occ(Steal)(3)$ is true can each be represented as follows:

$$\{In(1), In(2), In(3), Occ(Steal)(3), \neg In(4), \dots\}$$

Indeed, there is only one such world, namely w itself. By inertia, $At(3)$ and $Aff(At)(3)$ are true at w. The Law of Change and Inertia requires that $Aff(At)(3)$ has a cause, and so $\exists e\, Cause(Occ(e)(3), \bullet In(4))$ is true at w. Moreover, it follows from the definition of $SCause$ that $e = Steal$. So $Cause(Occ(Steal)(3), \neg In(4))$ is true at w. So it follows from the definition of $Expl$ that $Expl(Occ(Steal)(3), \neg In(4))$ is true at w.

The closest worlds to w at which $Occ(Steal)(2)$ is true can each be represented as follows:

$$\{In(1), In(2), Occ(Steal)(2), \neg In(3), \neg In(4), \dots\}$$

Clearly $Cause(Occ(Steal)(2), \neg In(3))$ is true at these worlds, and clearly $\bullet \exists e\, Causes(Occ(e)(3), \bullet\neg In(4))$ is also true. So $Expl(Occ(Steal)(2), \neg In(4))$ is true at w.

Similarly, the closest worlds to w at which $Occ(Steal)(1)$ is true can each be represented as follows:

$$\{In(1), Occ(Steal)(1), \neg In(2), \neg In(3), \neg In(4), \dots\}$$

Clearly $Cause(Occ(Steal)(1), \neg In(2))$ is true at these worlds, and clearly $\bullet \exists e, t(1 < t < 4 \wedge Causes(Occ(e)(t), \bullet\neg In(4))$ is also true. So $Expl(Occ(Steal)(1), \neg In(4))$ is true at w. ☐

8 Related work

The theory of primary events began with the work of Shoham [42, 43]. The language $\mathcal{TC}$ is based on his language TK. This was extended to QTK in [5], in order to allow for quantification and domain axioms, transferred to the three-valued base in [6], and extended to include its higher-order Situation Calculus-like features in [7, 50]. This makes it possible to express the general change and inertia axioms of Θ_P. As noted, these were inspired by those proposed by McCarthy [28]. They were also influenced by those proposed by Lifschitz [21]. As suggested, the pragmatics of causal theories began with Shoham's chronological minimization. The divergence begins with the refinement to prioritized chronological minimization in order that inertia and qualifications can both be represented in the same theory; examples 14 and 18.

Secondary events and most of the ideas presented here were first proposed in an earlier version of this paper [8].

The use of secondary events in order to represent ramifications has something in common with Thielscher's treatment of ramifications. Thielscher starts with a STRIPS-like representation of events. The ramifications of an event are then brought about by applying a series of causal constraints (causally directed domain axioms) until a stable state is reached. Although conceptually very different from the theory proposed in Section 4, the general effect in appropriate cases is similar; for example, the chain of invoked secondary events in Example 31 is similar to the successive application of causal constraints (although of course the former does not involve a sequence of intermediary states). And, indeed, it is clear from examples such as Example 31, that it is possible to represent Thielscher's three circuit examples correctly using primary and secondary events. However, as Thielscher starts with STRIPS-like events, his approach cannot, as it stands, deal with qualifications, non-determinism or simultaneous events. Sandewall's Causal Propagation Semantics [40] is very similar to Thielscher's approach, and has similar limitations.

Secondary events should be compared with the natural actions proposed by Lin, Pinto and Reiter [24, 33, 34]. In addition to actions initiated by agents, they suggest that there are also natural actions which arise due to the nature of the world (the Laws of Nature, etc.), and which are guaranteed to succeed. Their use is illustrated by Pinto's treatment of Lifschitz's lamp circuit example; the first circuit discussed by Thielscher. If the switches in a circuit are in the same position (both up or both down) then a lamp is on, otherwise it is off. Initially the switches are in opposite positions and one of them is flipped. Clearly we expect the lamp to be on as a result. However attempts to formalize this example using domain axioms with an inertia axiom fail. In the intended models the other switch retains its position by inertia and so the lamp comes on as a result of the flip event.

However there are also unintended models in which the lamp remains off by inertia and consequently the other switch mysteriously changes position as a result; this problem is essentially the same as that posed by Example 28. In Pinto's treatment of this example, the agent flipping a switch results in the natural action of current flowing in the circuit, and this in turn results in the lamp being on. The natural action of the current flowing resolves the conflict between alternative uses of the inertia axiom by bringing about an intermediary state in which both switches are up (as the flowing current does not affect their position) and in which the inertia axiom can only be applied in the intended way (as the flowing current is guaranteed to have the effect that the lamp is on). The underlying idea thus seems to be to ensure the causal directedness of domain axioms. However, as natural actions are guaranteed to succeed, it is difficult to see how they could be used to represent qualifications; for example, if the agent flips the switch then the current might not flow if the wire loses its conductivity, is cut, etc. By contrast, it is clear how to represent these using secondary events; in a context in which the switches are in opposite positions, flipping a switch invokes a secondary event which, if it succeeds, results in the lamp being on; the lamp's coming on is thus a defeasible secondary effect of flipping the switch. Secondary events thus differ from natural actions in that while the former are guaranteed to succeed the latter are not; recall that secondary events are invoked by primary events, that their success is dependent on the success of the primary event or events which invoke them, and that they may fail when the invoking event or events succeed. Moreover, while natural actions bring about an intermediary state, and so succeed the agent's action, secondary effects are contemporaneous with the primary events which invoke them.

Several formal theories of causal reasoning employ a primitive causes predicate; for example, an unanalyzed causes predicate is used in the $\mathcal{A}$-language family originating with Gelfond and Lifschitz [14] and in Lin's extension of the Situation Calculus which is discussed below. This begs the question. If the causes predicate represents a substantial notion of causation, then what *is* that notion? Alternatively, if it is merely syntactic sugar then its use appears ad hoc. The substantial difference in this paper is that causation is defined in the terms of the theory, and the definition is motivated by philosophical considerations on the nature of causality.

The Law of Change and Inertia has some similarity with the frame axiom proposed by Lin [23]. He adds a ternary *Caused* predicate to the Toronto dialect of the Situation Calculus [25]. Thus $Caused(p, v, s)$ "is true if the fluent p is caused (by something unspecified) to have the truth value v in the situation s" [23, p. 1986]. In keeping with this interpretation, two axioms are given:

$$Caused(p, true, s) \supset Holds(p, s)$$
$$Caused(p, false, s) \supset \neg Holds(p, s)$$

Thus if fluent p is caused to have the value true (false) in situation s, then p holds (does not hold) in situation s. By way of illustration, Lin discusses the example of a spring loaded suitcase. If both of its locks are up, the suitcase opens. Initially the suitcase is closed, one lock is up the other is down, nd the down-lock is flipped. Natually we expect that the suitcase will open as a result. However, attempts to formalize this example using an ordinary domain axiom and inertia fail; as in the lamp circuit example, the other lock remains up in the intended models, but there are also unintended models in which the suitcase remains closed and the other lock changes position as a result. Consequently Lin proposes a causally directed domain axiom:

$$up(L1, s) \wedge up(L2, s) \supset Caused(open, true, s) \tag{108}$$

Thus if locks $L1$ and $L2$ are both up in situation s, then the suitcase is caused to be open in s. His frame axiom is as follows:

$$\begin{aligned} Poss(a, s) \rightarrow \{\neg(\exists v)\, Caused(p, v, do(a, s)) \supset \\ [Holds(p, do(a, s)) \equiv Holds(p, s)]\}. \end{aligned}$$

So, if action a is executable in situation s and fluent p is not caused to have the value v in the situation resulting from performing a in s, then the truth value of p is unchanged by performing a in s, and so it persists to the resulting situation $do(a, s)$. Thus: "Unless caused otherwise, a fluent's truth value will persist" [p. 1986]. The predicate *Caused* is then circumscribed in the appropriate set of axioms, with the effect that:

$$Caused(open, v, s) \equiv v = true \wedge up(L1, s) \wedge up(L2, s)$$

Thus the suitcase is open in situation s if and only if both $L1$ and $L2$ are up in s. So the fact that both locks are up in s causes the suitcase to be open in s. While there is some similarity between Lin's frame axiom and the Law of Change and Inertia, it hardly goes beyond the level of the maxim that nothing changes without a cause. According to Lin's account, ramifications arise as a result of causally directed domain axioms such as (108), in which facts cause other facts to change. Indeed, Lin says of (108) "that the physical spring loaded mechanism behind the causal rule has been abstracted away. For all we care, it may just as well be that the device is not made of spring, but of bombs that will blow open the suitcase each time the two locks are in the up position. It then seems natural to say that the fluent *open* is caused to be true by the fact that the two locks are both in the up position" [p. 1986]. This seems odd from a common sense point of view, which has it that events bring about change and that facts are otherwise inert. Moreover, as the actual cause has been abstracted away, it is difficult to see how Lin's account could be extended in order to include the treatment of qualifications; for example, the suitcase may fail to open when both locks are up because the mechanism is rusty, someone is sitting on the suitcase, etc. By contrast, it is obvious how to represent the problem using primary and secondary events; in a context in which one lock is up and the other is down, flipping the down-lock invokes a secondary event which, if it succeeds, has the additional effect that the suitcase is open. Finally, in addition to regulating inertia, the law of Change and Inertia also regulates events by forcing the failure of those whose success would give rise to uncaused effects thereby ensuring a balance between change and inertia; Example 43.

The idea of using theories of causal reasoning such as the one given in [5] in order to solve the pragmatic problem for causal counterfactuals was first suggested in my thesis [4], and the idea of exploring alternative models/worlds above the preferred models/worlds of a causal theory in order to obtain explanations was suggested in [7]. The developments in this paper provide a much improved basis on which to resume the exploration of these ideas.

9 Concluding remarks

This paper has developed a fairly comprehensive theory of events and actual causal reasoning, and indicated how it can be extended to counterfactual causal reasoning. The theory can be extended in many other ways. For example, it is straightforward to remove the further STRIPS assumption that events are instantaneous. In order to do so, time points can be replaced by intervals (pairs of time points), thus the sentence $Occ(e)(i)$ could be used to state that event e occurs in the interval $i = [t, t']$; where t and t' are time points; this is done, for example, in [5, 6, 10, 43]. The current theory is also easily extended to complex events as in [10, p. 233-234]; for example, the sentence $Occ(e; e')(i)$ could be used to state that the complex event consisting of the occurrence of e followed by the occurrence of e' occurred during the interval i. Moreover, as the representation of time is based on that of temporal logic, it is readily generalized. In particular, time can be taken to be continuous, and so the present theory is readily extended to include the representation of continuous change by employing methods such as those in [37, 38, 41]. Finally, as illustrated in Section 7, $\mathcal{TC}$ is readily embedded in partial modal semantics. Thus it is readily integrated with work which uses classical possible-worlds semantics. This will be done with the theories of perception, belief and obligation presented in [9, 10] as part of a larger theory of practical reasoning and rationality. Before doing so however, the theory of causation given in Section 6 and the account of counterfactuals given in Section 7 will be developed further.

Acknowledgements

I am grateful to everyone who has commented on this work. I would particularly like to thank Sergio Brandano, Wilfrid Hodges, Fangzhen Lin, Erik Sandewall, Murray Shanahan, Michael Thielscher and Graham White for helpful discussions. This work was in part supported by the United Kingdom Engineering and Physical Sciences Research Council under grants GR/K19266 and GR/L34914.

References

[1] A. Avron (1988) Foundations of proof theory for 3-valued logics. Technical report number ECS-LFCS-88-84, Department of Computer Science, University of Edinburgh.

[2] A. Baker (1989) A simple solution to the Yale Shooting Problem. Proc. KR'89, pp. 11-20.

[3] A. Baker (1991) Nonmonotonic Reasoning in the Framework of the Situation Calculus. *Artificial Intelligence* 49, pp. 5-23.

[4] J. Bell (1988) *Predictive conditionals, nonmonotonicity and reasoning about the future*. Ph.D. thesis, University of Essex.

[5] J. Bell (1991) Extended causal theories. *Artificial Intelligence* 48, pp. 211-224.

[6] J. Bell (1996) A model-based approach to predictive causal reasoning. In [12], pp. 169-195.

[7] J. Bell (1998) Chronological Minimisation and Explanation. In [30], pp. 330-348.

[8] J. Bell (1999) Primary and Secondary Events. Earlier version of this paper. In M. Thielscher (ed.), Working notes of the 4th workshop on *Nonmonotonic Reasoning, Action and Change*, IJCAI'99, Stockholm, pp. 65-72.

[9] J. Bell and Z. Huang (1999) Seeing is believing: a common sense theory of the adoption of perception-based beliefs. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 13, pp. 133-140.

[10] J. Bell and Z. Huang (1999) Dynamic obligation hierarchies. In: *Norms, Logics and Information Systems*, P. McNamara and H. Prakken (eds.). Ios Press, Amsterdam.

[11] S. Blackburn (1994) *The Oxford Dictionary of Philosophy.* Oxford University Press, Oxford.

[12] P. Doherty (ed.) (1996) *Partiality, Modality, and Non-Monotonicity.* FoLLI Studies in Logic Language and Information series, CSLI Publications, Stanford.

[13] R.E. Fikes and N.J. Nilsson (1971) STRIPS; a new approach to the application of theorem proving to problem solving, *Artificial Intelligence* 2, pp. 189-208.

[14] M. Gelfond and V. Lifschitz (1993) Representing action and change by logic programs. *Journal of Logic Programming*, 17, pp. 301-322.

[15] S. Hanks and D. McDermott (1986) Default reasoning, nonmonotonic logics, and the frame problem. Proc. AAAI'86, pp. 328-333.

[16] S. Hanks and D. McDermott (1987) Nonmonotonic logic and temporal projection. *Artificial Intelligence* 33, pp. 379-412.

[17] D. Hume (1888) *A Treatise of Human Nature.* L.A. Selby-Bigge (ed.). Oxford University Press. Oxford.

[18] H.A. Kautz (1986) The logic of persistence, Proc. AAAI'86, pp. 401-405.

[19] S.C. Kleene (1952) *Introduction to Metamathematics.* North-Holland, Amsterdam.

[20] D. Lewis (1973) *Counterfactuals.* Basil Blackwell, Oxford.

[21] V. Lifschitz (1987) Formal theories of action. In F.M. Brown (ed.), *The Frame Problem in Artificial Intelligence*. Morgan Kaufmann, San Francisco, pp. 35-57.

[22] V. Lifschitz (1990) Frames in the Space of Situations, *Artificial Intelligence* 46, pp. 365-376.

[23] F. Lin (1995) Embracing causality in specifying the indirect effects of actions. Proc. IJCAI'95, pp. 1985-1993.

[24] F. Lin (1998) On the Relationships between Static and Dynamic Causal Rules in the Situation Calculus, Working Papers of the AAAI'98 Spring Symposium on *Prospects for a Commonsense Theory of Causation*, C. Ortiz (ed.), pp. 38-43.

[25] F. Lin and R. Reiter (1994) State constraints revisited. *Journal on Logic and Computation, special issue on actions and processes* 4(5) pp. 655-678.

[26] J.L. Mackie (1975) Causes and Conditions. In [44], pp. 15-38. First published in American Philosophical Quarterly 2.4, 1965, pp. 245-255 and 261-264.

[27] N. McCain and H. Turner (1995) A causal theory of ramifications and qualifications. Proc. IJCAI'95, pp. 1978-1984.

[28] J. McCarthy (1986) Applications of Circumscription to formalizing common sense knowledge, *Artificial Intelligence* 26, pp. 89-116.

[29] J. McCarthy and P. Hayes (1969) Some philosophical problems from the standpoint of Artificial Intelligence. In: *Machine Intelligence 4*, D. Michie and B. Meltzer (eds.), Edinburgh University Press, Edinburgh.

[30] R. Miller and M. Shanahan (eds.) (1998) Working papers of the *4th Symposium on Logical Formalizations of Commonsense Reasoning*, London. Available at http://www.ida.liu.se/ext/etai/nj/fcs-98/listing.html.

[31] N.J. Nilsson (1980) *Principles of Artificial Intelligence*, Addison Wesley, Menlo Park, California.

[32] R. Penrose (1989) *The Emperor's New Mind*. Oxford University Press, Oxford.

[33] J. Pinto (1998) Causality in Theories of Action. In [30], pp. 349-364.

[34] R. Reiter (1996) Natural Actions, Concurrency and Continuous Time in the Situation Calculus, Proc. KR'96, pp. 2-13.

[35] N. Rescher and A. Urquhart (1971) *Temporal Logic.* Springer, Berlin.

[36] J. Rosser and A. Turquette (1952) *Many-valued Logics.* North-Holland, Amsterdam.

[37] E. Sandewall (1989) Combining logic and differential equations for describing real-world systems. Proc. KR'89, pp. 412-420.

[38] E. Sandewall (1989) Filter preferential entailment for the logic of action in almost continuous worlds. Proc. IJCAI'89, pp. 894-899.

[39] E. Sandewall (1994) *Features and Fluents*, Oxford University Press, Oxford.

[40] E. Sandewall (1996) Assessments of ramification methods that use static domain constraints. Proc. KR'96, pp. 99-110.

[41] M. Shanahan (1997) *Solving the Frame Problem.* M.I.T. Press, Cambridge, Mass.

[42] Y. Shoham (1986) Chronological ignorance: time, nonmonotonicity, necessity and causal theories. Proc. AAAI'86, pp. 389-393.

[43] Y. Shoham (1988) *Reasoning About Change*, M.I.T. Press, Cambridge Mass.

[44] E. Sosa (1975) *Causation and Conditionals*, Oxford University Press, Oxford, 1975.

[45] R. Stalnaker (1968) A theory of conditionals. In N. Rescher (ed.) *Studies in Logical Theory*, Basil Blackwell, Oxford. Reprinted in [44].

[46] M. Thielscher (1997) Ramification and causality, *Artificial Intelligence* 98, pp. 317-364.

[47] J. van Benthem (1986) Partiality and Nonmonotonicity in Classical Logic. *Logique et Analyse* 29, pp. 225-247.

[48] J. van Benthem (1991) *The Logic of Time*, 2nd edition. Kluwer academic Publishers, Dordrecht.

[49] F. Veltman (1985) *Logics for Conditionals.* Ph.D. thesis, University of Amsterdam.

[50] G. White, J. Bell and W. Hodges (1998) Building Models of Prediction Theories. Proc. 6th Intl. Conference on the Principles of Knowledge Representation and Reasoning, A. Cohn, L. Schubert and S. Shapiro (eds.), Morgan Kaufmann, San Francisco, pp. 557-568.