

Some Logical Properties of Nonmonotonic Causal Theories

Marek Sergot and Robert Craven

Department of Computing, Imperial College London, SW7 2AZ, UK
{mjs,rac101}@doc.ic.ac.uk

Abstract. The formalism of nonmonotonic causal theories (Giunchiglia, Lee, Lifschitz, McCain, Turner, 2004) provides a general-purpose formalism for nonmonotonic reasoning and knowledge representation, as well as a higher level, special-purpose notation, the action language $\mathcal{C}+$, for specifying and reasoning about the effects of actions and the persistence (‘inertia’) of facts over time. In this paper we investigate some logical properties of these formalisms. There are two motivations. From the technical point of view, we seek to gain additional insights into the properties of the languages when viewed as a species of conditional logic. From the practical point of view, we are seeking to find conditions under which two different causal theories, or two different action descriptions in $\mathcal{C}+$, can be said to be equivalent, with the further aim of helping to decide between alternative formulations when constructing practical applications.

1 Introduction

The formalism of nonmonotonic causal theories, presented by Giunchiglia, Lee, Lifschitz, McCain and Turner [1], is a general-purpose language for knowledge representation and nonmonotonic reasoning. A *causal theory* is a set of causal rules each of which is an expression of the form

$$F \Leftarrow G$$

where F and G are formulas of an underlying propositional language and $F \Leftarrow G$ corresponds to the statement “if G , then F has a cause” (which is not the same as saying that G is a cause for F).

Associated with causal theories is the action language $\mathcal{C}+$, also presented in [1]. This may be viewed as a higher-level formalism for defining classes of causal theories in a concise and natural way, for the purposes of specifying and reasoning about the effects of actions and the persistence, or ‘inertia’, of facts through time, with support for indirect effects, non-deterministic actions and concurrency. The two closely-related formalisms have been used to represent standard domains from the knowledge representation literature.

In this paper, we investigate some logical properties of these formalisms. There are two motivations. The first is technical, to gain new insights into the

languages when they are viewed as species of conditional logic. For example, Turner [2] presents a more general formalism called the ‘logic of universal causation’. A rule $F \Leftarrow G$ of a causal theory can be expressed equivalently in this logic by the formula

$$G \rightarrow \mathbf{C} F$$

where \mathbf{C} is a modal operator standing for ‘there is a cause for’ (and \rightarrow is truth-functional, ‘material’ implication). Since \mathbf{C} is a normal modal operator whose logic is at least as strong as $S5$, some logical properties of $F \Leftarrow G$ are immediately obvious. For example, we can see from $G \rightarrow \mathbf{C} F \vdash_{S5} G \rightarrow \mathbf{C}(F \vee H)$ that the logic of causal theories will exhibit the property of ‘weakening of the consequent’: $F \Leftarrow G$ implies (in a sense to be made more precise) $(F \vee H) \Leftarrow G$. Other properties of $F \Leftarrow G$ will be straightforwardly propositional, such as ‘strengthening of the antecedent’: $F \Leftarrow G$ implies $F \Leftarrow G \wedge H$. This last property is intriguing, since it is often seen as a characteristic feature of monotonic conditionals, yet the logic of causal theories is nonmonotonic.

In this paper, we will not rely on the translation to Turner’s modal logic but prove properties directly from the semantics of causal theories. This is largely because of space limitations. Although many of the properties can be derived quite straightforwardly in $S5$, there is some preliminary notation and terminology we need, and we do not have space to introduce it here. Moreover, there are some fundamental properties of causal theories that are not inherited from $S5$.

The second motivation is a practical one. Causal theories and $\mathcal{C}+$ are very expressive languages. One purpose of the technical investigation is to find conditions under which two different causal theories, or two different action descriptions in $\mathcal{C}+$, can be said to be equivalent, with the further aim of helping to decide between alternative formulations when constructing applications.

2 Causal theories

A *multi-valued propositional signature* σ [3, 1] is a set of symbols called *constants*. For each constant c in σ , there is a non-empty set $\text{dom}(c)$ of values called the *domain* of c . An *atom* of a signature σ is an expression of the form $c=v$, where c is a constant in σ and $v \in \text{dom}(c)$. A *formula* φ of signature σ is any propositional compound of atoms of σ .

A *Boolean constant* is one whose domain is the set of truth values $\{\mathbf{t}, \mathbf{f}\}$. If p is a Boolean constant, p is shorthand for the atom $p=\mathbf{t}$ and $\neg p$ for the atom $p=\mathbf{f}$. Notice that, as defined here, $\neg p$ is an *atom* when p is a Boolean constant.

An *interpretation* of σ is a function that maps every constant in σ to an element of its domain. An interpretation I *satisfies* an atom $c=v$, written $I \models c=v$, if $I(c) = v$. The satisfaction relation \models is extended from atoms to formulas in accordance with the standard truth tables for the propositional connectives. When X is a set of formulas we also write $I \models X$ to signify that $I \models \varphi$ for all formulas $\varphi \in X$. I is then a *model* for the set of formulas X .

We write $\models_{\sigma} \varphi$ to mean that $I \models_{\sigma} \varphi$ for all interpretations I of σ . Where X is a set of formulas of signature σ , $X \models_{\sigma} \varphi$ denotes that $I \models_{\sigma} \varphi$ for all

interpretations I of σ such that $I \models_{\sigma} X$. When X' is a set of formulas of signature σ , $X \models_{\sigma} X'$ is shorthand for $X \models_{\sigma} \varphi$ for all formulas $\varphi \in X'$. In addition, where A and B are sets of formulas of a multi-valued propositional signature, we define $A \equiv_{\sigma} B$ to mean that $A \models_{\sigma} B$ and $B \models_{\sigma} A$. A *causal rule* is an expression of the form $F \Leftarrow G$, where F and G are formulas of signature σ . A *causal theory* is a set of causal rules.

Semantics Let Γ be a causal theory, and let X be an interpretation of its underlying propositional signature. Then the *reduct* of Γ , written Γ^X , is

$$\{F \mid F \Leftarrow G \in \Gamma \text{ and } X \models G\}$$

X is a model of Γ , written $X \models_{\mathcal{C}} \Gamma$, iff X is the unique model of the reduct Γ^X . By $\text{models}(\Gamma)$ we denote the set of all models of the causal theory Γ .

Γ^X is the set of all formulas which have a cause to be true, according to the rules of Γ , under the interpretation X . If Γ^X has no models, or has more than one model, or if it has a unique model different from X , then X is not considered to be a model of Γ . Γ is *consistent* or *satisfiable* iff it has a model.

For an illustration of the preceding definitions, consider the causal theory T_1 , with underlying Boolean signature $\{p, q\}$: $T_1 = \{p \Leftarrow q, q \Leftarrow q, \neg q \Leftarrow \neg q\}$.

There are clearly four possible interpretations of the signature: $X_1(p \mapsto \mathbf{t}, q \mapsto \mathbf{t})$, $X_2(p \mapsto \mathbf{t}, q \mapsto \mathbf{f})$, $X_3(p \mapsto \mathbf{f}, q \mapsto \mathbf{t})$, and $X_4(p \mapsto \mathbf{f}, q \mapsto \mathbf{f})$. It is clear that $T_1^{X_1} = \{p, q\}$, whose only model is X_1 ; $T_1^{X_2} = \{\neg q\}$, which has two models; $T_1^{X_3} = \{p, q\}$, whose only model is $X_1 \neq X_3$; and $T_1^{X_4} = \{\neg q\}$, which has two models. In only one of these cases—that of X_1 —is it true that the reduct of the causal theory with respect to the interpretation has that interpretation as its unique model. Thus $X_1 \models_{\mathcal{C}} T_1$ and $\text{models}(T_1) = \{X_1\}$.

Suppose we add another law to T_1 : for example, $T_2 = T_1 \cup \{\neg p \Leftarrow \neg p\}$. Now we have $\text{models}(T_2) = \{X_1, X_4\}$. In this example, augmenting the causal theory *increases* the set of models. It is clear that in general, for causal theories Γ and Δ , $\text{models}(\Gamma \cup \Delta) \not\subseteq \text{models}(\Gamma)$. This is the sense in which the causal theories are nonmonotonic. In the following, one of our purposes will be to investigate under which conditions $\Gamma \cup \Delta$ has the same models as Γ .

3 A consequence relation between causal theories

In this section, we frequently omit set-theoretic brackets from causal theories where doing so does not create confusion. In particular, causal theories which are singletons are often represented by the sole law they contain.

Proposition 1. $X \models_{\mathcal{C}} \Gamma$ iff, for every formula F , $X \models F$ iff $\Gamma^X \models_{\sigma} F$.

Proof. This is Proposition 1 of [1]. □

Observation 2 $(\Gamma_1 \cup \Gamma_2)^X = \Gamma_1^X \cup \Gamma_2^X$.

Proposition 3. $X \models \{F \Leftarrow G\}^X$ iff $X \models G \rightarrow F$.

Proof. Assume $X \models \{F \Leftarrow G\}^X$. If $X \models G$, then $\{F \Leftarrow G\}^X = \{F\}$, so $X \models F$. For the other direction, suppose $X \models G \rightarrow F$. If $X \models G$ then $X \models F$. But then $\{F \Leftarrow G\}^X = \{F\}$ and we have $X \models \{F \Leftarrow G\}^X$. If $X \not\models G$ then $\{F \Leftarrow G\}^X = \emptyset$, and $X \models \{F \Leftarrow G\}^X$, trivially. \square

It follows from the above that $X \models \{F_1 \Leftarrow G_1, \dots, F_n \Leftarrow G_n\}^X$ iff $X \models (G_1 \rightarrow F_1) \wedge \dots \wedge (G_n \rightarrow F_n)$. Moreover, if a causal theory Γ contains a rule $F \Leftarrow G$ then every model of Γ satisfies $G \rightarrow F$, i.e., $X \models_{\mathcal{C}} \Gamma$ implies $X \models G \rightarrow F$. This last remark is Proposition 2 of [1].

Where Γ is a causal theory, we will denote by $mat(\Gamma)$ the set of formulas obtained by replacing every rule $F \Leftarrow G$ of Γ by the corresponding material implication, $G \rightarrow F$. The remarks above can thus be summarised as follows.

Proposition 4.

$$(i) X \models \Gamma^X \text{ iff } X \models mat(\Gamma) \quad (ii) X \models_{\mathcal{C}} \Gamma \text{ implies } X \models mat(\Gamma)$$

Proof. In the preceding discussion. \square

We now define a notion of consequence between causal theories. This will allow us to say under which conditions two causal theories are equivalent, to simplify causal theories by removing causal laws that are implied by the causal theory, and to identify (in the following section) general properties of causal laws.

We will say that causal theories Γ_1 and Γ_2 of signature σ are equivalent, written $\Gamma_1 \equiv \Gamma_2$, when $\Delta \cup \Gamma_1$ and $\Delta \cup \Gamma_2$ have the same models for all causal theories Δ of signature σ . We will say that Γ_1 implies Γ_2 , written $\Gamma_1 \vdash \Gamma_2$, when $(\Gamma_1 \cup \Gamma_2) \equiv \Gamma_1$, that is, when $\Delta \cup \Gamma_1 \cup \Gamma_2$ and $\Delta \cup \Gamma_1$ have the same models for all causal theories Δ of signature σ .

Proposition 5. $\Gamma_1 \equiv \Gamma_2$ iff $\Gamma_1 \vdash \Gamma_2$ and $\Gamma_2 \vdash \Gamma_1$

Proof. A straightforward consequence of the definitions and basic set theory. \square

Proposition 6. $\Gamma \vdash \Gamma_1$ and $\Gamma \vdash \Gamma_2$ iff $\Gamma \vdash (\Gamma_1 \cup \Gamma_2)$

Proof. Immediate from the definitions. \square

Proposition 7. For all causal theories $\Gamma, \Gamma_1, \Gamma_2, \Delta$ of signature σ we have:

- (i) If $\Gamma_1 \equiv \Gamma_2$, then $(\Gamma_1 \cup \Delta) \vdash \Gamma$ iff $(\Gamma_2 \cup \Delta) \vdash \Gamma$.
- (ii) If $\Gamma_1 \equiv \Gamma_2$, then $\Gamma \vdash (\Delta \cup \Gamma_1)$ iff $\Gamma \vdash (\Delta \cup \Gamma_2)$.
- (iii) If $\Gamma_1 \equiv \Gamma_2$, then $(\Gamma_1 \cup \Delta) \equiv (\Gamma_2 \cup \Delta)$.

Proof. Part (ii): suppose $\Gamma_1 \equiv \Gamma_2$ and $(\Gamma_1 \cup \Delta) \vdash \Gamma$. That $models(\Delta' \cup (\Gamma_2 \cup \Delta) \cup \Gamma)$ is equal to $models(\Delta' \cup (\Gamma_2 \cup \Delta))$ follows easily using basic set theory. The other parts can be proved in similar fashion. \square

Proposition 8. The relation \vdash between causal theories of a given signature σ is a classical consequence relation (also known as a closure operator), that is, it satisfies the following three properties, for all causal theories Γ, Γ' , and Δ :

- **inclusion:** $\Gamma \vdash \Gamma$
- **cut:** $(\Gamma \cup \Delta) \vdash \Gamma'$ and $\Gamma \vdash \Delta$ implies $\Gamma \vdash \Gamma'$
- **monotony:** $\Gamma \vdash \Gamma'$ implies $(\Gamma \cup \Delta) \vdash \Gamma'$

Proof. ‘Inclusion’ is trivial. For ‘monotony’, suppose $\Gamma \vdash \Gamma'$. Then $(\Gamma \cup \Gamma') \equiv \Gamma$. We show $(\Gamma \cup \Delta \cup \Gamma') \equiv (\Gamma \cup \Delta)$ for any causal theory Δ . Clearly $(\Gamma \cup \Delta \cup \Gamma') \equiv ((\Gamma \cup \Gamma') \cup \Delta)$. And $((\Gamma \cup \Gamma') \cup \Delta) \equiv (\Gamma \cup \Delta)$ because $(\Gamma \cup \Gamma') \equiv \Gamma$. The proof for ‘cut’ is similar. \square

Corollary 1.

$\Gamma_1 \vdash \Gamma_2$ iff $(\Gamma_1 \cup \Gamma_2) \equiv \Gamma_1$ (As $\Gamma_1 \vdash \Gamma_1$ and $\Gamma_1 \vdash \Gamma_2$ iff $\Gamma_1 \equiv (\Gamma_1 \cup \Gamma_2)$.)
If $\Gamma_1 \vdash \Gamma_2$ and $\Gamma_2 \vdash \Gamma_3$ then $\Gamma_1 \vdash \Gamma_3$. (By ‘monotony’ and ‘cut’.)

Notice that although the formalism of causal theories is non-monotonic, in the sense that in general $\text{models}(\Gamma \cup \Delta) \not\subseteq \text{models}(\Gamma)$, the consequence relation \vdash between causal theories is monotonic.

We now establish some simple *sufficient* conditions under which \vdash holds.

Proposition 9. $\text{models}(\Gamma_1) \subseteq \text{models}(\Gamma_2)$ iff, for all $X \in \text{models}(\Gamma_1)$, we have $\Gamma_1^X \equiv_\sigma \Gamma_2^X$.

Proof. Omitted. \square

Corollary 2. $\text{models}(\Gamma_1) = \text{models}(\Gamma_2)$ iff we have $\Gamma_1^X \equiv_\sigma \Gamma_2^X$, for all $X \in \text{models}(\Gamma_1) \cup \text{models}(\Gamma_2)$.

Proposition 10.

- (i) $\Gamma_1 \vdash \Gamma_2$ if $\Gamma_1^X \models_\sigma \Gamma_2^X$, for all $X \models \text{mat}(\Gamma_1 \cup \Gamma_2)$.
- (ii) $\Gamma_1 \vdash \Gamma_2$ if $\Gamma_1^X \models_\sigma \Gamma_2^X$, for all $X \models \text{mat}(\Gamma_1)$.
- (iii) $\Gamma \vdash (G \Leftarrow F)$ if $\Gamma^X \models_\sigma G$, for all $X \models \text{mat}(\Gamma_1) \cup \{F\}$.

Proof. Part (i) follows from considering Proposition 9 and Corollary 2; the details of this have been omitted. Part (ii) is obtained from Part (i) by strengthening the condition. Part (iii) follows from Part (ii): if $X \models F$ then $\{G \Leftarrow F\}^X = \{G\}$. If $X \not\models F$ then $\{G \Leftarrow F\}^X = \emptyset$, and so $\Gamma^X \models_\sigma \{G \Leftarrow F\}^X$, trivially. \square

We record one further property for future reference. A causal rule of the form $F \Leftarrow F$ expresses that F holds by default. Adding $F \Leftarrow F$ to a causal theory Γ cannot eliminate models, though it can add to them.

Proposition 11. $\text{models}(\Gamma) \subseteq \text{models}(\Gamma \cup \{F \Leftarrow F\})$

Proof. Suppose $X \models_c \Gamma$, i.e., $X \models \Gamma^X$ and $Y \models \Gamma^X$ implies $Y = X$. We show (i) $X \models (\Gamma \cup \{F \Leftarrow F\})^X$, and (ii) if $Y \models (\Gamma \cup \{F \Leftarrow F\})^X$ then $Y = X$. For (i): if $X \models F$ then $(\Gamma \cup \{F \Leftarrow F\})^X = \Gamma^X \cup \{F\}$; we have both $X \models \Gamma^X$ and $X \models F$. If $X \not\models F$, then $(\Gamma \cup \{F \Leftarrow F\})^X = \Gamma^X$; we have $X \models \Gamma^X$. For (ii), suppose $Y \models (\Gamma \cup \{F \Leftarrow F\})^X$. Then $Y \models \Gamma^X$, and $X \models_c \Gamma$ implies $Y = X$. \square

4 Properties of \Leftarrow

We can now prove properties about the logic of causal theories, using the preliminary results and definitions given in the previous section. We have chosen to name the results after Chellas's [4] taxonomy of rules of inference from modal logic, as this scheme is well-known and seems natural to us. We frequently omit the proofs in this section, which in nearly all cases are based straightforwardly on Proposition 10, possibly including reasoning by cases of interpretations. (Full proofs are available in a companion technical report.)

In the following, we will frequently use the notational convenience of writing $\frac{A}{B}$ instead of $A \vdash B$, where A and B are causal rules or sets of such.

Proposition 12. [RCM] If $F_1 \models_{\sigma} F_2$, then $F_1 \Leftarrow G \vdash F_2 \Leftarrow G$

Proof. From Proposition 10(iii), a sufficient condition for $F_1 \Leftarrow G \vdash F_2 \Leftarrow G$ is $\{F_1 \Leftarrow G\}^X \models_{\sigma} F_2$ for all $X \models_{\sigma} G$, which is just $F_1 \models_{\sigma} F_2$, which was given. \square

Proposition 13. [RAug] If $G_1 \models_{\sigma} G_2$, then $F \Leftarrow G_2 \vdash F \Leftarrow G_1$

Proof. Similar to that for Proposition 12, and also using Proposition 10(ii). \square

Given the preceding two propositions and Proposition 5, we have the following corollary, of which the second part will be dubbed [RCEA], again after [4].

Corollary 3.

$$\begin{array}{l} (i) \text{ If } F_1 \equiv_{\sigma} F_2, \text{ then } F_1 \Leftarrow G \equiv F_2 \Leftarrow G \\ \text{[RCEA]} \quad (ii) \text{ If } G_1 \equiv_{\sigma} G_2, \text{ then } F \Leftarrow G_1 \equiv F \Leftarrow G_2 \end{array}$$

Proposition 14.

$$\text{[RCK]} \quad \text{If } F_1, \dots, F_n \models_{\sigma} F, \text{ then } \frac{F_1 \Leftarrow G, \dots, F_n \Leftarrow G}{F \Leftarrow G} \quad (n \geq 0)$$

The above are properties characteristic of 'normal conditional logics' [4]. We now move on to consider some distribution laws.

Proposition 15.

$$\begin{array}{ll} \text{[CC]} & \frac{F_1 \Leftarrow G, \dots, F_n \Leftarrow G}{(F_1 \wedge \dots \wedge F_n) \Leftarrow G} & \text{[CM]} & \frac{(F_1 \wedge \dots \wedge F_n) \Leftarrow G}{F_1 \Leftarrow G, \dots, F_n \Leftarrow G} \\ \text{[DIL]} & \frac{F \Leftarrow G_1, \dots, F \Leftarrow G_n}{F \Leftarrow (G_1 \vee \dots \vee G_n)} & \text{[cDIL]} & \frac{F \Leftarrow (G_1 \vee \dots \vee G_n)}{F \Leftarrow G_1, \dots, F \Leftarrow G_n} \end{array}$$

There follows a network of interrelated properties which all express a form of monotonicity of the conditional \Leftarrow .

Proposition 16. [Aug] $F \Leftarrow G \vdash F \Leftarrow G \wedge H$

Proof. This is clearly a specific instance of [RAug]. For a direct proof: a sufficient condition for [Aug] is $\{F \Leftarrow G\}^X \models_\sigma F$ for all $X \models (G \wedge H)$. But if $X \models (G \wedge H)$, then the condition reduces to $F \models_\sigma F$, which holds. \square

In fact, it can be shown that in the presence of the rule [RCEA], which we proved as Corollary 3(ii), the schema [Aug] is equivalent to the distribution law [cDIL]; constraints on space prevent our including the proof.

Proposition 17. [Contra] $\vdash F \Leftarrow \perp$

Proof. A sufficient condition for this is $\emptyset^X \vdash F$ for all $X \models \perp$, which holds trivially, since there is no such X . \square

Proposition 18. $F \Leftarrow G \vdash \perp \Leftarrow \neg F \wedge G$

Proof. By Proposition 10(iii), a sufficient condition is that $\{F \Leftarrow G\}^X \models_\sigma \perp$, for all X with $X \models (\neg F \wedge G) \wedge (G \rightarrow F)$, which obtains: there is no such X . \square

The converse of this proposition does not hold: $\perp \Leftarrow \neg F \wedge G \not\vdash F \Leftarrow G$. Now, since $G \models_\sigma \perp$ iff $G \equiv_\sigma \perp$, the schema [Contra] is equivalent to the rule: if $G \models_\sigma \perp$ then $\vdash F \Leftarrow G$. This is the case $n = 0$ for the following generalization of [RAug], easily proved by induction on n .

Proposition 19.

$$[\text{RDIL}] \quad \text{If } (G_1 \vee \dots \vee G_n) \models_\sigma G, \text{ then } \frac{F \Leftarrow G_1, \dots, F \Leftarrow G_n}{F \Leftarrow G} \quad (n \geq 0)$$

Proposition 20. [S] $F \Leftarrow G, G \Leftarrow H \vdash F \Leftarrow H$

A statement of the propagation of constraints, and a rule of Modus Ponens, are obvious instances of [S]:

Corollary 4.

$$[\text{Constr}] \quad \frac{F \Leftarrow G, \quad \perp \Leftarrow F}{\perp \Leftarrow G}, \quad [\text{MP}] \quad \frac{F \Leftarrow G, \quad G \Leftarrow \top}{F \Leftarrow \top}$$

From $G \Leftarrow \top$, we get $\perp \Leftarrow \neg G \wedge \top$ using Proposition 18; an application of [RCEA] then gives us $\perp \Leftarrow \neg G$. Using [Contra] and [S] we then derive $F \Leftarrow \neg G$:

Proposition 21. $G \Leftarrow \top \vdash F \Leftarrow \neg G$

The rule describing propagation of constraints may be generalised to a weak resolution law for Horn-like rules:

Proposition 22. $F \Leftarrow G \wedge H, G \Leftarrow K \vdash F \Leftarrow H \wedge K$

The logic of causal theories does *not* contain the two equivalent rules

$$[\text{I}] \quad \vdash F \Leftarrow F; \quad [\text{RI}] \quad \text{If } G \models_\sigma F, \text{ then } \vdash F \Leftarrow G$$

To see this, use $\neg p \Leftarrow \neg p$ for F and consider the causal theory with the single rule $p \Leftarrow p$; $\text{models}(\{p \Leftarrow p\}) \neq \text{models}(\{p \Leftarrow p, \neg p \Leftarrow \neg p\})$ which means that we do not have $\text{models}(\Gamma \cup \emptyset) = \text{models}(\Gamma \cup \emptyset \cup \{\neg p \Leftarrow \neg p\})$ for all causal theories Γ , and so $\not\vdash \neg p \Leftarrow \neg p$. Although we do not have [I], it has already been shown (Proposition 11) that $\text{models}(\Gamma) \subseteq \text{models}(\Gamma \cup \{F \Leftarrow F\})$.

Using the same example, it can easily be seen that the logic of \Leftarrow does not contain a contrapositive law: we have that $F \Leftarrow G \not\vdash \neg G \Leftarrow \neg F$.

Example As one example of an application of these properties consider the following common patterns of causal rules:

$$\{F \Leftarrow F \wedge G \wedge \neg R, \neg F \Leftarrow R\} \quad \text{and} \quad \{F \Leftarrow F \wedge G, \neg F \Leftarrow R\}$$

In each case the first law expresses that F holds by default if G holds, and the second that R is an exception to the default rule. These pairs of laws are equivalent in causal theories. One direction is straightforward: $F \Leftarrow F \wedge G \wedge \neg R$ follows from $F \Leftarrow F \wedge G$ by [Aug].

For the other direction, notice first that $\neg F \Leftarrow R$ implies $\perp \Leftarrow F \wedge G \wedge R$ (because $\neg F \Leftarrow R$ implies $\perp \Leftarrow F \wedge R$ and the rest follows by [Aug]). Now $\vdash F \Leftarrow \perp$ [Contra], and by [S] we derive $F \Leftarrow F \wedge G \wedge R$. For the final step

$$\frac{F \Leftarrow F \wedge G \wedge R, \quad F \Leftarrow F \wedge G \wedge \neg R}{F \Leftarrow F \wedge G \wedge (R \vee \neg R)}$$

from which $F \Leftarrow F \wedge G$ follows.

5 The action language $\mathcal{C}+$

As with the logic of causal theories, the language $\mathcal{C}+$ is based on a multi-valued propositional signature σ , with σ partitioned into a set σ^f of *fluent constants* and a set σ^a of *action constants*. Further, the fluent constants are partitioned into those which are *simple* and those which are *statically determined*. A *fluent formula* is a formula whose constants all belong to σ^f ; an *action formula* has at least one action constant and no fluent constants.

A *static law* is an expression of the form

$$\text{caused } F \text{ if } G,$$

where F and G are fluent formulas. An *action dynamic law* is an expression of the same form in which F is an action formula and G is a formula. A *fluent dynamic law* has the form

$$\text{caused } F \text{ if } G \text{ after } H,$$

where F and G are fluent formulas and H is a formula, with the restriction that F must not contain statically determined fluents. *Causal laws* are static laws or dynamic laws, and an *action description* is a set of causal laws.

In the following section we will make use of several of the many abbreviations afforded in $\mathcal{C}+$. In particular:

α *causes* F if G abbreviates the fluent dynamic law **caused** F if \top **after** $\alpha \wedge G$;
nonexecutable α if G expresses that there is no transition of type α from a state satisfying fluent formula G . It is shorthand for the fluent dynamic law **caused** \perp if \top **after** $\alpha \wedge G$;

inertial f where f is a simple fluent constant, states that the values of f persist by default—they are subject to inertia—from one state to the next. It stands for the collection of fluent dynamic laws **caused** $f=v$ if $f=v$ **after** $f=v$ for every $v \in \text{dom}(f)$.

exogenous a where a is an action constant, stands for the set of action dynamic laws **caused** $a=v$ if $a=v$ **after** $a=v$ for every $v \in \text{dom}(a)$.

The language $\mathcal{C}+$ can be viewed as a useful shorthand for the logic of causal theories, for to every action description D of $\mathcal{C}+$ and non-negative integer m , there corresponds a causal theory Γ_m^D . The signature of Γ_m^D contains constants $c[i]$, such that

- $i \in \{0, \dots, m\}$ and c is a fluent constant of the signature of D , or
- $i \in \{0, \dots, m-1\}$ and c is an action constant of the signature of D ,

and the domains of such constants $c[i]$ are kept identical to those of their constituents c . The expression $F[i]$, where F is a formula, denotes the result of suffixing $[i]$ to every occurrence of a constant in F . The causal rules of Γ_m^D are: $F[i] \Leftarrow G[i]$, for every static law in D and every $i \in \{0, \dots, m\}$, and for every action dynamic law in D and every $i \in \{0, \dots, m-1\}$; $F[i+1] \Leftarrow G[i+1] \wedge H[i]$, for every fluent dynamic law in D and every $i \in \{0, \dots, m-1\}$; and $f[0]=v \Leftarrow f[0]=v$, for every simple fluent constant f and $v \in \text{dom}(c)$.

Each action description of $\mathcal{C}+$ defines a labelled transition system. The states of the transition system are the models of Γ_0^D . Each state is an interpretation of the fluent constants σ^f . Transitions are the models of Γ_1^D . Each transition label (also called an ‘event’) is an interpretation of the action constants σ^a . When α is a formula of σ^a , we say that a transition label/event is of type α when it satisfies the formula α . It can also be shown [1] that paths of length m of the transition system correspond to the models of Γ_m^D .

Since action descriptions are a shorthand for particular forms of causal theories, we obtain a notion of equivalence and consequence between action descriptions of $\mathcal{C}+$: $D_1 \vdash_{\mathcal{C}+} D_2$ is defined as $\Gamma_m^{D_1} \vdash \Gamma_m^{D_2}$ for all non-negative integers m , and likewise for equivalence of action descriptions. We also inherit properties of causal laws of $\mathcal{C}+$ corresponding to those proven in Section 4. For example: α **causes** F if $G \vdash_{\mathcal{C}+} \alpha \wedge \beta$ **causes** F if G follows directly from ‘augmentation’ [Aug] for causal theories in general. In the following we will use the same labels for properties of $\mathcal{C}+$ as for the causal-theoretic properties from which they are derived.

6 Example (Winning the lottery)

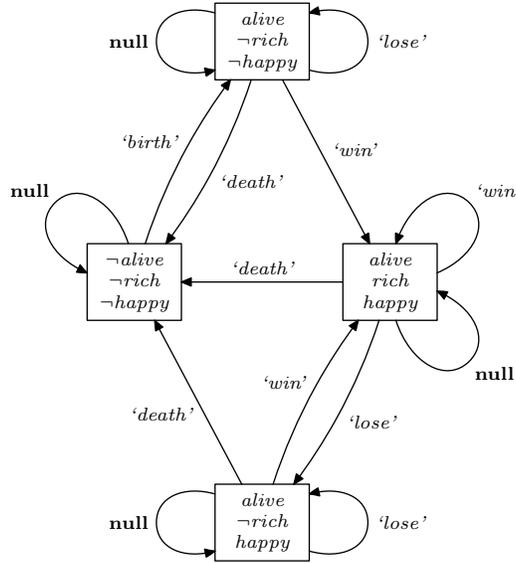
Winning the lottery causes one to become (or remain) rich. Losing one’s wallet causes one to become (or remain) not rich. A person who is rich is happy. A person who is not alive is neither rich nor happy.

The example is constructed partly to show how $\mathcal{C}+$ deals with indirect effects of actions (ramifications). The example also illustrates some issues in the

representation of concurrent actions, actions with defeasible effects, and non-deterministic actions. Naturally it is not possible to illustrate everything with one simple example, but the example is indicative of the issues that are encountered when formulating applications in a language as expressive as $\mathcal{C}+$.

Signature: simple Boolean fluent constants *alive, rich, happy*; Boolean action constants *birth, death, win, lose*.

inertial *alive, rich, happy*
 exogenous *birth, death, win, lose*
birth causes *alive*
 nonexecutable *birth* if *alive*
death causes \neg *alive*
 nonexecutable *death* if \neg *alive*
win causes *rich*
 nonexecutable *win* if \neg *alive*
lose causes \neg *rich*
 nonexecutable *lose* if \neg *alive*
 caused *happy* if *rich*
 caused \neg *rich* if \neg *alive*
 caused \neg *happy* if \neg *alive*
 nonexecutable *birth* \wedge *death*
 nonexecutable *birth* \wedge *win*
 nonexecutable *birth* \wedge *lose*
 nonexecutable *win* \wedge *lose*



States and transition labels/events are interpretations of the fluent constants and action constants, respectively. Here, each state and each transition label/event is represented by the set of atoms that it satisfies. Because of the static laws, there are only four states in the transition system and not $2^3 = 8$. The diagram label *'birth'* is shorthand for the label/event $\{birth, \neg death, \neg win, \neg lose\}$, and likewise for the labels *'death'*, *'win'* and *'lose'*. The label **null** is shorthand for $\{\neg birth, \neg death, \neg win, \neg lose\}$. The diagram does not show transitions of type *death* \wedge *lose*, *win* \wedge *death*, and so on. We will discuss those presently.

Notice that *happy* is declared inertial, and so still persists even if one becomes not rich. That is why the *'lose'* transition from state $\{alive, rich, happy\}$ results in the state $\{alive, \neg rich, happy\}$. We could of course modify the action description so that *happy* is no longer inertial but defined to be true if and only if *rich* is true. Or we might prefer to make *happy* non-inertial and let the *'lose'* transition be non-deterministic. The interactions between these various adjustments are rather subtle, however, and are not always immediately obvious.

We will restrict attention to the following two questions. First, there are alternative ways of formulating the constraints that a person cannot be rich or happy when not alive, and these alternatives have different interactions with the other causal laws. Second, as it turns out, the last group of four nonexecutable

statements are all redundant, in that they are already implied by the other causal laws. There are some remaining questions about the effects of concurrent actions in the example which we will seek to identify.

First, let us look at some effects of individual actions. With the static constraints as formulated above, we have the following implied laws. (Henceforth we omit the keyword *caused* to conserve space.) *death causes* \neg *alive* (in other words, \neg *alive* if \top after *death*) together with \neg *rich* if \neg *alive* imply *death causes* \neg *rich*. And in general

$$\frac{\alpha \text{ causes } F \text{ if } G, \quad F' \text{ if } F}{\alpha \text{ causes } F' \text{ if } G}$$

as is easily checked. By a similar argument we also have the implied causal law *death causes* \neg *happy* (\neg *happy* if \top after *death*) and *win causes* *happy*. We do not get the law *lose causes* \neg *happy* because as formulated here, we do not have the static law (explicit or implied) \neg *happy* if \neg *rich*.

Suppose that instead of the static laws \neg *rich* if \neg *alive* and \neg *happy* if \neg *alive*, we had included only the weaker constraints \perp if *rich* \wedge \neg *alive* and \perp if *happy* \wedge \neg *alive*. These constraints eliminate the unwanted states, but are too weak to give the implied effects (ramifications). We also lose transitions: if \neg *happy* if \neg *alive* is replaced by either of *alive* if *happy* or \perp if *happy* \wedge \neg *alive*, the only way that \neg *happy* can be ‘caused’ is by inertia. Consequently, we eliminate all the *death* transitions from states in which *happy* holds: we get the implied law nonexecutable *death* if *happy*. (We omit the formal derivation of this implied law for lack of space. It is rather involved since it also requires taking into account the presence of other causal laws in the example.) Similarly, if we replace \neg *rich* if \neg *alive* by either of *alive* if *rich* or \perp if *rich* \wedge \neg *alive*, the only way that \neg *rich* can be ‘caused’ is by a *lose* transition or by inertia. Consequently, transitions of type *death* \wedge \neg *lose* become non-executable in the states $\{alive, rich, \neg happy\}$ and $\{alive, rich, happy\}$ whether or not we also make the earlier adjustment to the *alive/happy* constraint. In addition, we have the implied law nonexecutable *death* \wedge \neg *lose* if *rich*: a rich person cannot die unless he simultaneously loses his wallet.

There is one way in which we can use constraints \perp if *rich* \wedge \neg *alive* and \perp if *happy* \wedge \neg *alive* (or *alive* if *rich* and *alive* if *happy*) without losing transitions. That is by adding a pair of extra fluent dynamic laws: either

$$death \text{ causes } \neg rich \quad \text{and} \quad death \text{ causes } \neg happy$$

or the weaker pair *death may cause* \neg *rich* and *death may cause* \neg *happy*. (In $C+$, α may cause F is an abbreviation for the fluent dynamic law F if F after α .) We leave out the (straightforward) derivation that demonstrates both these pairs have the claimed effect. Neither is entirely satisfactory since they require all ramifications of *death* to be identified in advance and then modelled explicitly using causal laws.

We turn now to examine the effects of concurrent actions. First, notice that the law nonexecutable *birth* \wedge *death* is implied by the other causal laws. Because: *alive* if \top after *birth* and \neg *alive* if \top after *death* imply by [Aug]

alive if \top after *birth* \wedge *death* and \neg *alive* if \top after *birth* \wedge *death*, which in turn together imply by [CC] *alive* \wedge \neg *alive* if \top after *birth* \wedge *death* (which is equivalent to nonexecutable *birth* \wedge *death*). And in general

$$\frac{\alpha \text{ causes } A \text{ if } F, \quad \beta \text{ causes } B \text{ if } G, \quad C \text{ if } A \wedge B}{\alpha \wedge \beta \text{ causes } C \text{ if } F \wedge G}$$

There is another derivation of nonexecutable *birth* \wedge *death* from the causal laws of the example. We have the causal laws nonexecutable *birth* if *alive* and nonexecutable *death* if \neg *alive*. \perp if \top after *birth* \wedge *alive* and \perp if \top after *death* \wedge \neg *alive* imply by [Aug]: \perp if \top after *birth* \wedge *death* \wedge *alive* and \perp if \top after *birth* \wedge *death* \wedge \neg *alive*, which in turn together imply by [DIL] \perp if \top after (*birth* \wedge *death* \wedge *alive*) \vee (*birth* \wedge *death* \wedge \neg *alive*), whose antecedent can be simplified by [RCEA]: \perp if \top after *birth* \wedge *death*.

In general we have:

$$\frac{\text{nonexecutable } \alpha \text{ if } F, \quad \text{nonexecutable } \beta \text{ if } G}{\text{nonexecutable } \alpha \wedge \beta \text{ if } (F \vee G)}$$

What of *birth* \wedge *win* and *birth* \wedge *lose*? We have

$$\frac{\text{nonexecutable } \textit{birth} \text{ if } \textit{alive}, \quad \text{nonexecutable } \textit{win} \text{ if } \neg \textit{alive}}{\text{nonexecutable } \textit{birth} \wedge \textit{win}}$$

from which nonexecutable *birth* \wedge *lose* follows by a similar argument.

This leaves transitions of type *death* \wedge *lose* and *death* \wedge *win*. *death* \wedge *lose* is not problematic. We have the implied causal laws *death* \wedge *lose* causes \neg *alive* (by [Aug] from *death* causes \neg *alive*) and *death* \wedge *lose* causes \neg *rich* (either by [Aug] from *lose* causes \neg *rich* or from the implied law *death* causes \neg *rich*). In this example, the effects of *death* \wedge *lose* transitions are the same as those of *death* \wedge \neg *lose* transitions.

Consider now *death* \wedge *win*. Here we need some adjustment to the example's formulation. We have the implied law nonexecutable *win* \wedge *death* because (one of several possible derivations): we have the implied law (*win* \wedge *death*) causes (*rich* \wedge \neg *alive*), the static law \neg *rich* if \neg *alive* implies \perp if *rich* \wedge \neg *alive*, and so (*win* \wedge *death*) causes \perp , which is equivalent to nonexecutable *win* \wedge *death*.

But it seems unreasonable to insist that *win* \wedge *death* transitions cannot happen—that was not the intention when the example was originally formulated. We can admit the possibility of *win* \wedge *death* transitions by re-formulating the relevant causes statement for *win* so that it reads instead *win* causes *rich* if \neg *death*, or equivalently *win* \wedge \neg *death* causes *rich*. The effects of the 'win' transitions are unchanged, but the transition system now contains transitions of type *win* \wedge *death*: their effects are exactly the same as those of 'death' and *death* \wedge *lose* transitions.

But note that after this adjustment, we have to re-examine other combinations of possible concurrent actions. *win* \wedge *birth* is still non-executable (it depended on the pre-conditions of the two actions, not their effects) but we no longer have nonexecutable *win* \wedge *lose*. We have only the implied law (*win* \wedge

lose) causes (*rich* \wedge \neg *rich*) if \neg *death*, or equivalently, nonexecutable *win* \wedge *lose* \wedge \neg *death*. So now a person can win the lottery and lose his wallet simultaneously, but only if he dies at the same time.

But suppose *win* \wedge *lose* \wedge \neg *death* is intended to be executable. What should its effects be? One possibility is that the effects of *win* override those of *lose*. We replace the *lose* causes \neg *rich* law by the weaker *lose* causes \neg *rich* if \neg *win*. A second possibility is that the effects of *lose* override those of *win*. We replace the *win* causes *rich* if \neg *death* law by the weaker *win* causes *rich* if \neg *death* \wedge \neg *lose*. (And we may prefer to introduce an ‘abnormality’ action constant (see [1, Section 4.3]) to express the defeasibility of winning more concisely.) The third possibility is to say that *win* \wedge *lose* transitions are non-deterministic:

$$\textit{win} \wedge \textit{lose} \text{ may cause } \textit{rich}, \quad \textit{win} \wedge \textit{lose} \text{ may cause } \neg \textit{rich}$$

What of the interactions between non-deterministic *win* \wedge *lose* actions and *death*? We still have the implied law *win* \wedge *lose* \wedge *death* causes \neg *rich*. But perhaps the non-deterministic effects of the other *win* \wedge *lose* transitions should have been formulated thus:

$$\textit{win} \wedge \textit{lose} \wedge \neg \textit{death} \text{ may cause } \textit{rich}, \quad \textit{win} \wedge \textit{lose} \wedge \neg \textit{death} \text{ may cause } \neg \textit{rich}$$

This is unnecessary. In $\mathcal{C}+$ $\{\alpha \text{ may cause } F, \alpha \text{ may cause } \neg F, \beta \text{ causes } \neg F\}$ and $\{\alpha \wedge \neg \beta \text{ may cause } F, \alpha \wedge \neg \beta \text{ may cause } \neg F, \beta \text{ causes } \neg F\}$ are equivalent. Left-to-right is just an instance of [Aug]. For right-to-left, notice that $\alpha \wedge \neg \beta \text{ may cause } F, \beta \text{ causes } \neg F$ is an instance of the general pattern of causal rules $\{P \Leftarrow P \wedge Q \wedge \neg R, \neg P \Leftarrow R\}$, discussed at the end of Section 4. It is equivalent to $\{P \Leftarrow P \wedge Q, \neg P \Leftarrow R\}$. For the other part, notice that $\beta \text{ causes } \neg F$ implies $\alpha \wedge \beta \text{ may cause } \neg F$ by [Aug], and $\alpha \wedge \neg \beta \text{ may cause } \neg F$ and $\alpha \wedge \beta \text{ may cause } \neg F$ together imply $\alpha \text{ may cause } \neg F$ by [DIL] and [RCEA].

There are other variations of the example that we do not consider here for lack of space. We might remove the declaration that *happy* is inertial. Or we might choose to make the fluent constant *happy* statically determined instead of ‘simple’. These changes have a further set of interactions with the other causal laws. Their effects can be analysed in similar fashion.

References

1. Giunchiglia, E., Lee, J., Lifschitz, V., McCain, N., Turner, H.: Nonmonotonic causal theories. *Artificial Intelligence* **153** (2004) 49–104
2. Turner, H.: A logic of universal causation. *Artificial Intelligence* **113** (1999) 87–123
3. Giunchiglia, E., Lee, J., Lifschitz, V., Turner, H.: Causal laws and multi-valued fluents. In: Proc. of the Fourth Workshop on Nonmonotonic Reasoning, Action, and Change, Seattle. (2001)
4. Chellas, B.F.: *Modal Logic—An Introduction*. Cambridge University Press (1980)