On Relating Causal Theories to Other Formalisms

Norman McCain and Hudson Turner
Department of Computer Sciences
University of Texas at Austin
Austin, TX 78712–1188, USA
{mccain,hudson}@cs.utexas.edu

Abstract

This paper explores mathematical relationships between the "causal theories" formalism recently introduced by the authors and several other (well-known) formalisms. More specifically, it relates causal theories to default logic and autoepistemic logic, and describes translations back and forth between causal theories and classical propositional logic. It also relates action representations in causal theories to two previous causality-based proposals, due to Lin and to the authors.

1 Introduction

Applications to reasoning about action have motivated much of the work on nonmonotonic formalisms. This is particularly true of the "causal theories" formalism discussed in this paper. This mathematically simple system, based on causality, is introduced in a companion paper [McCain and Turner, 1997] which emphasizes underlying motivations and applications to commonsense knowledge about actions. By contrast, the current, complementary paper concentrates on mathematical issues.

Much work has been done on establishing connections between different logical formalisms, and different approaches to representing actions in those formalisms. The current paper continues this direction of research, relating causal theories to two well-known nonmonotonic formalisms — default logic and autoepistemic logic — and to classical propositional logic. It also relates the representation of commonsense knowledge about actions in causal theories to two previous, closely related approaches [Lin, 1995; McCain and Turner, 1995].

The models of causal theories are interpretations, in the sense of propositional logic. They assign a truth value to every atom, and in this sense they are "complete." In contrast, default logic and autoepistemic logic are characterized by "models" that are, in general, roughly speaking, "incomplete." We will show that causal theories are mathematically equivalent to syntactically restricted subsets of default logic and autoepistemic logic, in the special case when we consider only "models" that are consistent and "complete."

Causal theories can also be translated into classical propositional logic. In one translation, applicable to a syntactically restricted class of causal theories, the resulting classical propositional theories are of comparable size. In another, more generally applicable, translation — obtained indirectly through a translation into second-order propositional logic — the resulting theories are exponentially larger. (This translation assumes only that the signature of the language and the causal theory itself are finite.) Conversely, there is a simple translation of classical propositional theories into causal theories, in which the resulting causal theories are of comparable size. These complementary embeddings show that the two formalisms are equivalent (when we restrict our consideration to finite signatures and theories).

The approach to representing actions in causal theories described in the companion paper [McCain and Turner, 1997] is closely related to a proposal by Lin [1995] which is based on a similar form of causal knowledge. We establish a precise sense in which the two proposals agree on the possible results of performing an action in a given state. We consider this question also in regard to the proposal in [McCain and Turner, 1995], which is based on causal knowledge of a slightly different form.

2 Preliminary Definitions

We begin with a language of propositional logic, which includes the zero-ary logical connectives *True* and *False*.[1] A *literal* is an atom or its negation. We will identify each interpretation with the set of literals true in it.

We write inference rules as expressions of the form

$$\frac{\phi}{\psi}$$

where ϕ and ψ are formulas. We often find it convenient to identify a formula ϕ with the inference rule

$$\frac{True}{\phi}.$$

Let R be a set of inference rules and Γ a set of formulas. We say Γ is *closed under* R if for all $\frac{\phi}{\psi} \in R$, if $\phi \in \Gamma$

[1] *True* and $\neg$*False* are tautologies in which no atoms occur.

then $\psi \in \Gamma$. By $Cn(R)$ we denote the least logically closed set of formulas that is closed under R. Notice that $Cn(\Gamma) = \{\phi : \Gamma \models \phi\}$. Notice also that the set of formulas true in an interpretation I is $Cn(I)$.

2.1 Default Logic

Default logic is due to Reiter [1980]. A *default rule* is an expression of the form

$$\frac{\alpha : \beta_1, \ldots, \beta_n}{\gamma}$$

where all of $\alpha, \beta_1, \ldots, \beta_n, \gamma$ are formulas ($n \geq 0$). Let r be such a default rule. By $pre(r)$ we denote α, by $just(r)$ we denote the set $\{\beta_1, \ldots, \beta_n\}$, and by $cons(r)$ we denote γ. If $pre(r) = True$ we often omit it, writing $\frac{: \beta_1, \ldots, \beta_n}{\gamma}$ instead. Such rules are called *prerequisite-free*. If $just(r)$ is empty, we often find it convenient to identify r with the inference rule $\frac{\alpha}{\gamma}$.

A *default theory* is a set of default rules. A default theory is *prerequisite-free* if all of its rules are. Let D be a default theory and let E be a set of formulas. By D^E we denote the following set of inference rules.

$$\left\{ \frac{pre(r)}{cons(r)} \; : \; r \in D \textit{ and for all } \beta \in just(r), \neg\beta \notin E \right\}$$

We say E is an *extension* of D if

$$E = Cn(D^E)\,.$$

We say E is *complete* if there is an interpretation I such that $E = Cn(I)$. We will be particularly interested in complete extensions of prerequisite-free default theories.

2.2 Autoepistemic Logic

Autoepistemic (AE) logic is due to Moore [1985]. Syntactically, it is a modal logic, with modal operator B.

We can take AE *structures* to be pairs (I, S), where I is an interpretation and S is a set of interpretations [Lifschitz and Schwarz, 1993]. We define the truth of a AE formula in a AE structure (I, S) as follows.

$$\begin{array}{lll} (I,S) \models p & \textit{iff} & p \in I, \;\; p \textit{ is an atom} \\ (I,S) \models \neg\phi & \textit{iff} & (I,S) \not\models \phi \\ (I,S) \models \phi \wedge \psi & \textit{iff} & (I,S) \models \phi \textit{ and } (I,S) \models \psi \\ (I,S) \models B\phi & \textit{iff} & \textit{for all } I' \in S, \, (I',S) \models \phi \end{array}$$

Let T be an autoepistemic theory. A set S of interpretations is an *AE model* of T if

$$S = \{\, I : (I, S) \models T \,\}\,.$$

We say that S is *complete* if there is an interpretation I such that $S = \{I\}$. We will be particularly interested in complete AE models of autoepistemic theories.

3 Causal Theories

We begin with brief informal motivation. (For a more adequate account of the intuitions underlying causal theories, please see the companion paper.)

Intuitively, a "causally possible" world history is one that conforms to the true causal laws, i.e., one in which every fact that is caused (according to the true causal laws) obtains. We strengthen this idea by assuming the *principle of universal causation*, according to which every fact that obtains is caused.[2] Thus, we can say that a causally possible world history is one in which exactly the facts that obtain are caused.

Now assume that D is a complete description of the conditions under which facts are caused. In this case, we can say that a causally possible world history is one in which the facts that obtain are exactly those that are caused according to D. This is the key to understanding the formal definitions that follow. Notice that we make two assumptions: the principle of universal causation and the completeness of D.

3.1 Syntax

By a *causal law* we mean an expression of the form

$$\phi \Rightarrow \psi \tag{1}$$

where ϕ and ψ are formulas of the underlying propositional language. By the *antecedent* and *consequent* of (1), we mean the formulas ϕ and ψ, respectively. Note that (1) is not the material conditional $\phi \supset \psi$.

The intended reading of (1) is: *Necessarily, if ϕ then the fact that ψ is caused.* Often, but not always, we write (1) because we know something more, namely: *The fact that ϕ causes the fact that ψ.* The term "causal law" is suggested by this practice.

By a *causal theory* we mean a set of causal laws.

3.2 Semantics

For every causal theory D and interpretation I, let

$$D^I = \{\, \psi : \textit{for some } \phi,\; \phi \Rightarrow \psi \in D \textit{ and } I \models \phi \,\}\,.$$

That is, D^I is the set of consequents of all causal laws in D whose antecedents are true in I. Intuitively then, D^I entails exactly the formulas that are caused to be true in I according to D.

Main definition. Let D be a causal theory, and I be an interpretation. We say that I is *causally explained* according to D if I is the unique model of D^I.

Intuitively, when D describes an action domain, the causally explained interpretations according to D correspond to the causally possible world histories.

We have the following alternative characterization. An interpretation I is causally explained according to D if and only if for every formula ϕ

$$I \models \phi \;\; \textit{iff} \;\; D^I \models \phi\,.$$

Thus, intuitively, I is causally explained according to D if and only if the formulas that are true in I are exactly

[2] This rather strong philosophical commitment is rewarded by mathematical simplicity in the main definition of causal theories. Moreover, in applications it is easily relaxed. (See the companion paper.)

the formulas that are caused to be true in I according to D. Notice that this condition can also be written as

$$Cn(I) = Cn(D^I)\,.$$

4 Causal Theories as Default Theories

We translate a causal theory D into a prerequisite-free default theory $d(D)$ as follows.

$$d(D) = \left\{ \frac{:\phi}{\psi} : \phi \Rightarrow \psi \in D \right\}$$

The following theorem shows that the causally explained interpretations according to causal theory D correspond to the complete extensions of default theory $d(D)$.

Theorem 4.1 *An interpretation I is causally explained according to a causal theory D if and only if $Cn(I)$ is an extension of the default theory $d(D)$.*

Proof. It is easy to verify that for any interpretation I

$$Cn\left(D^I\right) = Cn\left(d(D)^{Cn(I)}\right)$$

which is sufficient to establish the theorem. □

5 Causal Theories as Autoepistemic Theories

We translate a causal theory D into an autoepistemic theory $ae(D)$ as follows.

$$ae(D) = \{\, B\phi \supset \psi : \phi \Rightarrow \psi \in D \,\}$$

The following theorem shows that the causally explained interpretations according to causal theory D correspond to the complete AE models of AE theory $ae(D)$.

Theorem 5.1 *An interpretation I is causally explained according to a causal theory D if and only if $\{I\}$ is an AE model of the autoepistemic theory $ae(D)$.*

Proof. Notice that for any nonmodal formulas ϕ and ψ

$$(I', \{I\}) \models B\phi \supset \psi \quad \textit{iff} \quad I \not\models \phi \textit{ or } I' \models \psi\,.$$

It follows that

$$(I', \{I\}) \models ae(D) \quad \textit{iff} \quad I' \models D^I\,.$$

So $\{I\} = \{\, I' : (I', \{I\}) \models ae(D) \,\}$ if and only if I is the unique model of D^I. □

6 Causal Theories as Classical Theories

We specify two translations of causal theories into theories of classical propositional logic. The first translation, applicable to a syntactically restricted class of causal theories, is an elaboration of the Clark completion method [Clark, 1978]. The second, due to Vladimir Lifschitz [1997], is obtained indirectly by means of a translation into second-order propositional logic. We also specify a translation of classical propositional logic into causal theories.

6.1 Literal Completion

Let D be a causal theory in which (i) the consequent of every causal law is a literal, and (ii) every literal is the consequent of finitely many causal laws. By the *literal completion* of D we mean the classical propositional theory obtained as follows: For each literal L in the language of D, include the formula

$$L \equiv (\phi_1 \vee \cdots \vee \phi_n)$$

where $\phi_1, \ldots, \phi_n$ are the antecedents of the causal laws with consequent L. We call this formula the *characteristic formula* of L.

Theorem 6.1 *Let D be a causal theory that satisfies conditions (i) and (ii) above. The causally explained interpretations according to D are precisely the models of its literal completion.*

Proof. Due to the first restriction on the form of D, we know that I is causally explained according to D if and only if $I = D^I$. Assume $I = D^I$. Then for every $L \in I$

- there is a causal law $\phi \Rightarrow L$ in D s.t. $I \models \phi$, and
- there is no causal law $\phi \Rightarrow \overline{L}$ in D s.t. $I \models \phi$.

(We write $\overline{L}$ to denote the literal complementary to L.) It follows that for every $L \in I$

- I satisfies the characteristic formula of L, and
- I satisfies the characteristic formula of $\overline{L}$.

Hence I is a model of the literal completion of D.

Proof in the other direction is similar. □

6.2 Via Second-Order Propositional Logic

Let D be a finite causal theory whose underlying language has a finite signature $\{P_1, \ldots, P_n\}$. Let $\{p_1, \ldots, p_n\}$ be a corresponding set of propositional variables. For each causal law $\phi \Rightarrow \psi \in D$, let $c(\phi \Rightarrow \psi)$ stand for the formula of second-order propositional logic obtained by replacing $\Rightarrow$ by $\supset$ and also replacing each occurrence of each atom P_i in ψ by the corresponding propositional variable p_i. The translation $c(D)$ of causal theory D is the following sentence of second-order propositional logic.

$$\forall p_1, \ldots, p_n \left[\left(\bigwedge_{\phi \Rightarrow \psi \,\in\, D} c(\phi \Rightarrow \psi) \right) \equiv \left(\bigwedge_{1 \le i \le n} p_i \equiv P_i \right) \right]$$

Theorem 6.2 *[Lifschitz, 1997] Let D be a finite causal theory whose underlying language has a finite signature. The causally explained interpretations according to D are precisely the models of the second-order propositional sentence $c(D)$.*

The propositional variables in the sentence $c(D)$ can be eliminated, at the cost of an exponential increase in length, as follows. For each interpretation I, let $p(D, I)$ be the sentence of propositional logic obtained from

$$\left(\bigwedge_{\phi \Rightarrow \psi \,\in\, D} c(\phi \Rightarrow \psi) \right) \equiv \left(\bigwedge_{1 \le i \le n} p_i \equiv P_i \right)$$

by replacing each occurrence of each propositional variable p_i by *True* if $I \models P_i$ and by *False* otherwise.

Corollary 6.2.1 *Let D be a finite causal theory whose underlying language has a finite signature. The causally explained interpretations according to D are precisely the models of the propositional sentence*

$$\bigwedge_I p(D,I)\,.$$

Although the corollary follows from Theorem 6.2, the following observations help explain independently why it holds. For any interpretation I, $I \models p(D,I)$ iff $I \models D^I$. Moreover, for any $I' \neq I$, $I \models p(D,I')$ iff $I' \not\models D^I$. Hence $I \models \bigwedge_I p(D,I)$ iff I is the unique model of D^I.

6.3 Classical Theories as Causal Theories

There is an extremely simple embedding of classical propositional theories in causal theories. (We omit the easy proof.) Given a classical propositional theory Γ, let $ct(\Gamma)$ denote the following causal theory.

$$\{\, True \Rightarrow \phi \,:\, \phi \in \Gamma \,\} \cup \{\, L \Rightarrow L \,:\, L \textit{ is a literal}\}$$

Theorem 6.3 *Let Γ be a classical propositional theory. The models of Γ are precisely the causally explained interpretations according to the causal theory $ct(\Gamma)$.*

7 Comparisons with Previous Causal Approaches to Representing Actions

In the representations of commonsense knowledge about actions proposed in [Lin, 1995] and [McCain and Turner, 1995], the central difficulty is understood to be the definition of "possible next states" — the states that can possibly result from performing an action in a given state. The causal theories for representing actions described in the companion paper [McCain and Turner, 1997] reflect a somewhat different view of things. But for the purpose of comparison, we focus here on the question of defining possible next states.

Also for the purpose of comparison, we present a simplified account of Lin's proposal. Most notably, we suppress the role of the situation calculus, and do not consider non-propositional fluent and action symbols.

Begin with a nonempty set of fluent atoms. The *fluent language* is the language of propositional logic whose atoms are the fluent atoms. A *fluent formula* (*literal*) is a formula (literal) in the fluent language. A *state* is an interpretation of the fluent language.[3]

We will compare three definitions of possible next states, each based on the following three parameters.

- An *initial state* S in which the action is performed.
- An *explicit effect* E: a set of formulas caused to be true as a direct effect of the action.[4]
- *Background knowledge* C: a set of causal laws characterizing the causal relationships between fluents.

This comparison framework is modeled after the definitions in [McCain and Turner, 1995].

7.1 Possible Next States: Causal Theories

For initial state S, explicit effect E, and background knowledge C, let $\Pi_1(S,E,C)$ be the set consisting of all states that are causally explained by the causal theory

$$\{L \Rightarrow L : L \in S\} \cup \{\, True \Rightarrow \phi : \phi \in E\} \cup C\,.$$

Notice that for any state S' we have

$$\{L \Rightarrow L : L \in S\}^{S'} = S \cap S'\,.$$

Intuitively, $S \cap S'$ consists of precisely the fluent literals that persist when moving from S to S'. So this component of the causal theory says that whenever the truth of a fluent literal persists, it is ("trivially") caused.

Notice also that for any state S'

$$\{\, True \Rightarrow \phi : \phi \in E\}^{S'} = E\,.$$

So the second component of the causal theory simply says that the explicit effect is caused.

We see that $S' \in \Pi_1(S,E,C)$ if and only if S' is the unique model of $(S \cap S') \cup E \cup C^{S'}$. This condition can be broken into two parts, as follows.

Lemma 7.1 *For any initial state S, explicit effect E, and background knowledge C, a state S' belongs to $\Pi_1(S,E,C)$ if and only if*

- $S' \models E \cup C^{S'}$ *and*
- $S' \setminus S \subseteq Cn((S \cap S') \cup E \cup C^{S'})\,.$

7.2 Possible Next States: Lin's Approach

We describe a simplification of the proposal from [Lin, 1995] that is adequate for defining possible next states.

We will employ circumscription on a first-order theory in a many-sorted language with two sorts, *fluent* and *value*. We construct this language on the basis of our (propositional) fluent language. We need to assume that the fluent language has a finite signature $\{F_1,\ldots,F_n\}$. The fluent atoms $F_1,\ldots,F_n$ serve, in the first-order language, as the object constants of sort *fluent*. The object constants of sort *value* are $\top$ and $\bot$. We include axioms expressing domain closure and unique names assumptions for both sorts, as follows, where f is a variable of sort *fluent* and v is a variable of sort *value*.

$$\forall f(f = F_1 \vee \cdots \vee f = F_n) \qquad (2)$$

$$\textstyle\bigwedge_{1 \leq i < j \leq n} F_i \neq F_j \qquad (3)$$

$$\forall v(v = \top \not\equiv v = \bot) \qquad (4)$$

The first-order language also includes two predicates — *Holds* and *Caused* — whose arities and sorts are clear in light of the following two additional axioms, which say

[3] Intuitively, some interpretations may not correspond to possible states, but we ignore this complication.

[4] We ignore the possibility that the explicit effect depends on the state in which an action is performed.

that a fluent is true whenever it is caused to be true, and false whenever it is caused to be false.

$$\forall f(\mathit{Caused}(f, \top) \supset \mathit{Holds}(f)) \quad (5)$$

$$\forall f(\mathit{Caused}(f, \bot) \supset \neg \mathit{Holds}(f)) \quad (6)$$

Given a fluent formula ϕ, $\mathit{Holds}(\phi)$ stands for the formula obtained by replacing every occurrence of every fluent atom F in ϕ by $\mathit{Holds}(F)$. For every fluent atom F, $\mathit{Caused}(F)$ stands for $\mathit{Caused}(F, \top)$, and $\mathit{Caused}(\neg F)$ stands for $\mathit{Caused}(F, \bot)$.

We need to assume that the explicit effect E is a finite set of literals. We also need to assume that the background knowledge C is finite, and that the consequents of all causal laws in C are literals. Let $\mathit{th}(E, C)$ be the first-order theory obtained by adding to axioms (2)–(6) the translations of E and C, specified as follows.

- Translate every fluent literal L in E as $\mathit{Caused}(L)$.
- Translate every causal law $\phi \Rightarrow L$ in C as

$$\mathit{Holds}(\phi) \supset \mathit{Caused}(L) .$$

The complete translation $\mathit{th}(S, E, C)$ is obtained by circumscribing *Caused*, with *Holds* fixed, in $\mathit{th}(E, C)$, and adding the appropriate frame axioms to the result, as follows. For each fluent atom F, if $F \in S$, include the axiom

$$\neg \mathit{Holds}(F) \equiv \mathit{Caused}(F, \bot) \quad (7)$$

otherwise include the axiom

$$\mathit{Holds}(F) \equiv \mathit{Caused}(F, \top) . \quad (8)$$

Let $\Pi_2(S, E, C)$ denote the set of all states S' for which there is a model M of $\mathit{th}(S, E, C)$ such that

$$M \models \bigwedge_{L \in S'} \mathit{Holds}(L) .$$

Theorem 7.1 *Assume that the fluent language has a finite signature. Given an initial state S, a finite explicit effect E consisting solely of literals, and finite background knowledge C such that all causal laws in C have literal consequents,* $\Pi_2(S, E, C) = \Pi_1(S, E, C)$.

Proof. Axioms (2)–(4) guarantee that every model of $\mathit{th}(E, C)$ is isomorphic to a Herbrand model. Therefore we can restrict our attention to Herbrand interpretations and otherwise forget axioms (2)–(4) in what follows.

Given a state S' and a set X of fluent literals, let $M(S', X)$ denote the Herbrand interpretation that satisfies the following conditions, for all fluent atoms F.

- $M(S', X) \models \mathit{Holds}(F)$ *iff* $F \in S'$
- $M(S', X) \models \mathit{Caused}(F, \top)$ *iff* $F \in X$
- $M(S', X) \models \mathit{Caused}(F, \bot)$ *iff* $\neg F \in X$

Every Herbrand interpretation can be expressed in this manner. Notice that $M(S', X)$ satisfies axioms (5) and (6) iff $X \subseteq S'$. It is easy to verify that $M(S', X)$ satisfies the translations of E and C iff $E \cup C^{S'} \subseteq X$. Since the circumscriptive policy effectively minimizes X, for a fixed S', it is clear from the previous observations that every Herbrand model of $\mathit{th}(S, E, C)$ can be written in the form $M(S', E \cup C^{S'})$, where $S' \models E \cup C^{S'}$. Moreover, one easily checks that $M(S', E \cup C^{S'})$ satisfies the axioms of forms (7) and (8) iff $S' \setminus S \subseteq E \cup C^{S'}$. Thus, by Lemma 7.1, we have established that $M(S', E \cup C^{S'})$ is a model of $\mathit{th}(S, E, C)$ iff $S' \in \Pi_1(S, E, C)$. Finally, observe that $M(S', E \cup C^{S'}) \models \bigwedge_{L \in S'} \mathit{Holds}(L)$. □

7.3 Possible Next States: Inference Rules

The definition of possible next states from [McCain and Turner, 1995] uses a set R of inference rules — not a set C of causal laws — to represent background knowledge of the causal relations between fluents. In that paper, S' is a possible next state if and only if

$$\mathit{Cn}(S') = \mathit{Cn}((S \cap S') \cup E \cup R) .$$

This is similar to the condition defining possible next states in Section 7.1, which can be written as

$$\mathit{Cn}(S') = \mathit{Cn}((S \cap S') \cup E \cup C^{S'}) .$$

Intuitively, an inference rule $\frac{\phi}{\psi}$ in R expresses that in every state in which ϕ is caused to be true, ψ is caused to be true. The corresponding causal law $\phi \Rightarrow \psi$ expresses something stronger — in every state in which ϕ *is true*, ψ is caused to be true.

Let S be an initial state, and E an explicit effect. Let C be background knowledge (in the form of causal laws). Take

$$R(C) = \left\{ \frac{\phi}{\psi} : \phi \Rightarrow \psi \in C \right\}$$

and let $\Pi_3(S, E, C)$ be the set of states S' such that

$$\mathit{Cn}(S') = \mathit{Cn}\left((S \cap S') \cup E \cup R(C)\right) .$$

We will say that $E \cup C$ is *stratified* if there is a mapping from the set of fluent atoms to the ordinals such that

- for every formula ϕ that belongs to E or is the consequent of a causal law in C, all atoms that occur in ϕ are mapped to the same ordinal, and
- for every causal law $\phi \Rightarrow \psi$ in C, every atom that occurs in ψ is mapped to a greater ordinal than every atom that occurs in ϕ.

The following theorem shows that substituting inference rules for causal laws in background knowledge can only eliminate possible next states. Moreover, if the explicit effect and background knowledge are stratified, the substitution does not change the possible next states.

Theorem 7.2 *For any initial state S, explicit effect E, and background knowledge C, $\Pi_3(S, E, C)$ is a subset of $\Pi_1(S, E, C)$. Moreover, if $E \cup C$ is stratified, then $\Pi_3(S, E, C) = \Pi_1(S, E, C)$.*

Due to space constraints, we do not present a proof of this theorem. We outline a proof for the first claim, as follows. Let $X = \mathit{Cn}((S \cap S') \cup E \cup R(C))$ and $Y = \mathit{Cn}((S \cup S') \cup E \cup C^{S'})$. Assume $\mathit{Cn}(S') = X$. Show that

X is closed under $C^{S'}$. It follows that $Y \subseteq X$. Show that Y is closed under $R(C)$. It follows that $X \subseteq Y$. So $X = Y$ and thus $Cn(S') = Y$. We describe the elements of a proof for the second claim, as follows. Theorem 4.1 shows that S' belongs to $\Pi_1(S, E, C)$ iff $Cn(S')$ is an extension of the default theory

$$\left\{ \frac{:L}{L} : L \in S \right\} \cup E \cup \left\{ \frac{:\phi}{\psi} : \phi \Rightarrow \psi \in C \right\}.$$

On the other hand, a theorem in [Przymusinski and Turner, 1997] shows that S' belongs to $\Pi_3(S, E, C)$ iff $Cn(S')$ is an extension of the default theory

$$\left\{ \frac{:L}{L} : L \in S \right\} \cup E \cup R(C).$$

The Splitting Sequence Theorem from [Turner, 1996] can be used to show that these two default theories have the same complete extensions when $E \cup C$ is stratified.

Theorem 7.2 shows that several results established in [McCain and Turner, 1995] for $\Pi_3(S, E, C)$ hold for $\Pi_1(S, E, C)$ as well. Thus, causal laws of the forms $\mathit{True} \Rightarrow \psi$ and $\phi \Rightarrow \mathit{False}$ can be understood to correspond to "ramification constraints" and "qualification constraints" (respectively) in the sense of Lin and Reiter [1994]. In fact, given background knowledge C expressed as causal laws of the form $\mathit{True} \Rightarrow \psi$, and the corresponding set $B = \{\psi : \mathit{True} \Rightarrow \psi \in C\}$ of state constraints, the possible next states according to $\Pi_1(S, E, C)$ are precisely those obtained using the state constraints B under Winslett's classic definition [1988].

8 Concluding Remarks

We have established embeddings of causal theories in default logic and autoepistemic logic. Thus, causal theories can be understood as a mathematically simple specialization of these highly expressive nonmonotonic formalisms. We have also specified translations back and forth between causal theories and classical propositional logic. Thus, these formalisms are equivalent (assuming finite signatures and theories as in Theorem 6.2), although the size explosion in the general translation of causal theories into propositional logic supports the claim that causal theories provide a more convenient representation of causal knowledge of the kind that concerns us.

These mathematical relationships show that computational methods developed for these standard formalisms are applicable as well to causal theories.[5] For instance, this means that under certain restrictions on the action domain — guaranteeing a complete specification of the initial state of the world accompanied by a complete absence of nondeterminism — it is possible to do satisfiability planning, in the sense of Kautz and Selman [1992], on the basis of action descriptions in causal theories.

We have demonstrated that action representations in causal theories are closely related to two previous causal approaches, due to Lin and to the authors. We showed that, in Lin's proposal, sentences of the form $\mathit{Holds}(\phi) \supset \mathit{Caused}(L)$ correspond in straightforward fashion to causal laws $\phi \Rightarrow L$. We also showed that causal laws are closely related to the inference rules used previously by the authors. We note that the definition of stratification introduced in Section 7 can be extended to the causal theories for representing actions described in the companion paper, in order to investigate in that more general setting the relationship between the following two closely related forms of causal knowledge.

- In every world in which ϕ is true, ψ is caused.
- In every world in which ϕ is caused, ψ is caused.

[5] One easily modifies the translations to obtain only the complete extensions and AE models.

Acknowledgements

Many thanks to our advisor, Vladimir Lifschitz. This work was partially supported by NSF grant IRI-9306751.

References

[Clark, 1978] Keith Clark. Negation as failure. In Herve Gallaire and Jack Minker, editors, *Logic and Data Bases*, pages 293–322. Plenum Press, New York, 1978.

[Kautz and Selman, 1992] Henry Kautz and Bart Selman. Planning as satisfiability. In J. Lloyd, editor, *Proceedings of the 10th European Conference on Artificial Intelligence (ECAI 92)*, Vienna, Austria, 1992.

[Lifschitz and Schwarz, 1993] Vladimir Lifschitz and Grigori Schwarz. Extended logic programs as autoepistemic theories. In Luis Moniz Pereira and Anil Nerode, editors, *Logic Programming and Non-monotonic Reasoning: Proceedings of the Second Int'l Workshop*, pages 101–114, 1993.

[Lifschitz, 1997] Vladimir Lifschitz. A logic for causal reasoning. Submitted for publication, 1997.

[Lin and Reiter, 1994] Fangzhen Lin and Raymond Reiter. State constraints revisited. *Journal of Logic and Computation*, 4(5):655–678, 1994. Special Issue on Actions and Processes.

[Lin, 1995] Fangzhen Lin. Embracing causality in specifying the indirect effects of actions. In *Proc. of IJCAI-95*, pages 1985–1991, 1995.

[McCain and Turner, 1995] Norman McCain and Hudson Turner. A causal theory of ramifications and qualifications. In *Proc. of IJCAI-95*, pages 1978–1984, 1995.

[McCain and Turner, 1997] Norman McCain and Hudson Turner. Causal theories of action and change. Submitted for publication, 1997.

[Moore, 1985] Robert Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25(1):75–94, 1985.

[Przymusinski and Turner, 1997] Teodor Przymusinski and Hudson Turner. Update by means of inference rules. *Journal of Logic Programming*, 30(2):125–143, 1997.

[Reiter, 1980] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1,2):81–132, 1980.

[Turner, 1996] Hudson Turner. Splitting a default theory. In *Proc. of AAAI-96*, pages 645–651, 1996.

[Winslett, 1988] Marianne Winslett. Reasoning about action using a possible models approach. In *Proc. of AAAI-88*, pages 89–93, 1988.