Pearl's Causality In a Logical Setting

Alexander Bochman
Computer Science Department
Holon Institute of Technology, Israel
bochmana@hit.ac.il

Vladimir Lifschitz
Department of Computer Science
University of Texas at Austin
vl@cs.utexas.edu

Abstract

We provide a logical representation of Pearl's structural causal models in the framework of the causal calculus of McCain and Turner (1997) and its first-order generalization by Lifschitz. It will be shown that, under this representation, the nonmonotonic semantics of the causal calculus describes precisely the solutions of the structural equations (the causal worlds of a causal model), while the causal logic from Bochman (2004) is adequate for describing the behavior of causal models under interventions (forming submodels).

1 Introduction

In recent years we witnessed successful uses of causal reasoning in two largely unrelated areas, reasoning about action and change in AI, and reasoning in statistics, economics, cognitive and social sciences, based mainly on Judea Pearl's theory (Pearl 2000). A common assumption behind both these approaches is that reasoning in the relevant domains cannot be expressed in the plain language of (classical) logic, but requires the explicit use of causal concepts and a general language of causation. In both cases, the corresponding causal formalisms have provided working concepts of reasoning that have turned out to be essential for an adequate representation of the respective areas, as well as for broad correspondence with commonsense descriptions. Nevertheless, studies in these two areas so far used apparently different formalisms and pursued quite different objectives.

In this study we are going to show that these two causal formalisms are based on essentially the same understanding of causation. We will do this by demonstrating that the central notion of Pearl's theory, the notion of a structural causal model (which is based, in turn, on the notion of a structural equation) can be naturally represented in the causal calculus of (McCain and Turner 1997), and especially in the first-order generalization of the latter, introduced in (Lifschitz 1997). Moreover, this representation creates a powerful generalization of Pearl's formalism that relaxes many of its more specific restrictions (such as acyclicity and uniqueness of solutions). In addition, it allows us to clarify some of the issues and confusions related to the use of causal concepts in Pearl's theory, such as the distinction between a plain mathematical and a causal understanding of structural equations.

In sum, we believe that our representation provides all the necessary conditions for returning Pearl's theory of causal reasoning to logic, with all the expected benefits such a logical representation could provide for its analysis, generalization and further development.

2 Structural Equations and Causal Models

Pearl's account

According to (Pearl 2000, Chapter 7), a causal model is a triple $M = \langle U, V, F\rangle$ where

(i) U is a set of *background* (or *exogenous*) variables.

(ii) V is a set $\{V_1, V_2, \ldots, V_n\}$ of *endogenous* variables that are determined by variables in $U \cup V$.

(iii) F is a set of functions $\{f_1, f_2, \ldots, f_n\}$ such that each f_i is a mapping from $U \cup (V \backslash V_i)$ to V_i, and the entire set, F, forms a mapping from U to V.

Symbolically, Pearl continues, F is represented as a set of equations

$$v_i = f_i(pa_i, u_i) \quad i = 1, \ldots, n$$

where pa_i is any realization of the unique minimal set of variables PA_i in $V \backslash \{V_i\}$ (parents) sufficient for representing f_i, and similarly for $U_i \subseteq U$.

In Pearl's account, every instantiation $U = u$ of the exogenous variables determines a particular "causal world" of the causal model. Such worlds stand in one-to-one correspondence with the solutions to the above equations in the ordinary mathematical sense. However, structural equations also encode causal information in their very syntax by treating the variable on the left-hand side of $=$ as the effect and treating those on the right as causes. Accordingly, the equality signs in structural equations convey the asymmetrical relation of "is determined by". This causal reading does not affect the set of solutions of a causal model, but it plays a crucial role in determining the effect of external interventions and evaluation of counterfactual assertions with respect to such a model (Pearl 2012).

Each structural equation in a causal model is intended to represent a stable and autonomous physical mechanism, which means that it is conceivable to modify (or cancel) one such equation without changing the others. In Pearl's theory, this modularity plays an instrumental role in determining the

answers to three types of queries that can be asked with respect to a causal model: *predictions* (e.g., will the pavement be slippery if we find the sprinkler off?), *interventions* (will the pavement be slippery if we make sure that the sprinkler is off?), and *counterfactuals* (would the pavement be slippery had the sprinkler been off, given that the pavement is in fact not slippery and the sprinkler is on?).

The answers to prediction queries can be obtained using plain deductive inferences from a logical description of the causal worlds. However, in order to obtain answers to the intervention (action) and counterfactual queries, we have to consider what was termed by Pearl submodels of a given causal model. Given a particular instantiation x of a set of variables X from V, a submodel M_x of M is described as the causal model that is obtained from M by replacing its set of functions F by the following set:

$$F_x = \{f_i \mid V_i \notin X\} \cup \{X = x\}.$$

In other words, F_x is formed by deleting from F all functions f_i corresponding to members of set X and replacing them with the set of constant functions $X = x$. A submodel M_x can be viewed as a result of performing an action $do(X = x)$ on M that produces a minimal change required to make $X = x$ hold true under any u. This submodel is used in Pearl's theory for evaluating counterfactuals of the form "Had X been x, whether $Y = y$ would hold?"

We will now recast Pearl's ideas in a form convenient for our analysis, starting with the propositional case.

Propositional case

Propositional formulas are formed from propositional atoms and the logical constants $\mathbf{f}$, $\mathbf{t}$ using the classical connectives $\wedge$, $\vee$, $\neg$, and $\rightarrow$.

Definition 1. Assume that the set of propositional atoms is partitioned into a set of of *background* (or *exogenous*) atoms and a finite set of *explainable* (or *endogenous*) atoms.

- A *Boolean structural equation* is an expression of the form $A = F$, where A is an endogenous atom and F is a propositional formula in which A does not appear.
- A *Boolean causal model* is a set of Boolean structural equations $A = F$, one for each endogenous atom A.

Definition 2. A *solution* (or a *causal world*) of a Boolean causal model M is any propositional interpretation satisfying the equivalences $A \leftrightarrow F$ for all equations $A = F$ in M.

Example 1. In the 'firing squad' example from (Pearl 2000, Chapter 7), let U, C, A, B, D stand, respectively, for the following propositions: "Court orders the execution", "Captain gives a signal", "Rifleman A shoots", "Rifleman B shoots", and "Prisoner dies." The story is formalized using the following causal model M, in which U is the only exogenous atom:

$$\{C = U,\ A = C,\ B = C,\ D = A \vee B\}.$$

It has two solutions: in one of them all atoms are true, in the other all atoms are false. This causal model allows us to answer 'static' queries concerning the domain. For instance, M implies $\neg A \rightarrow \neg D$, in the sense that this implication is satisfied by every causal world of M. That gives us a *prediction*:

S1. If rifleman A did not shoot, the prisoner is alive.

It implies also the implication $\neg D \rightarrow \neg C$, which amounts to *abduction*:

S2. If the prisoner is alive, the Captain did not signal.

And it implies $A \rightarrow B$, which amounts to *transduction*:

S3. If rifleman A shot, then B shot as well.

Definition 3. Given a Boolean causal model M, a subset X of the set of endogenous atoms, and a truth-valued function I on X, the *submodel* M^I_X of M is the causal model obtained from M by replacing every equation $A = F$, where $A \in X$, with $A = I(A)$.

Example 1, continued. Consider the action sentence

S4. If the captain gave no signal and rifleman A decides to shoot, the prisoner will die and B will not shoot.

To evaluate it, we need to consider the submodel $M^I_{\{A\}}$ of M with $I(A) = \mathbf{t}$:

$$\{C = U,\ A = \mathbf{t},\ B = C,\ D = A \vee B\}.$$

Since this submodel implies both $\neg C \rightarrow D$ and $\neg C \rightarrow \neg B$, S4 is justified.

First-order case

Terms are formed from object constants and function symbols as usual in first-order logic.

Definition 4. Assume that the set of object constants is partitioned into a set of *rigid* symbols, a set of of *background* (or *exogenous*) symbols, and a finite set of *explainable* (or *endogenous*) symbols.

- A *structural equation* is an expression of the form $c = t$, where c is an endogenous symbol and t is a ground term in which c does not appear.
- A *causal model* consists of an interpretation of the set of rigid symbols and function symbols (in the sense of first-order logic) and a set of structural equations $c = t$, one for each endogenous symbol c.

Definition 5. A *solution* (or a *causal world*) of a causal model M is an extension of the interpretation of rigid symbols and function symbols in M to the exogenous and endogenous symbols that satisfies all equalities $c = t$ in M.

Definition 6. Given a causal model M, a subset X of the set of endogenous symbols, and a function I from X to the set of rigid constants, the *submodel* M^I_X of M is the causal model obtained from M by replacing every equation $c = t$, where $c \in X$, with $c = I(c)$.

Example 2. Let us consider a closed gas container with variable volume that can be heated. P, V, T will denote, respectively, pressure, volume and temperature of the gas. In this particular setup, it is natural to treat P and V as endogenous,

while T as an exogenous symbol. The corresponding causal model will involve two structural equations that are actually two directional instances of the Ideal Gas Law:

$$P = c \cdot \frac{T}{V} \qquad V = c \cdot \frac{T}{P}$$

where c is a (rigid) constant. The language may include also names of real numbers; we classify them as rigid. A causal model is constructed by combining the above equations with the interpretation that has the set of positive real numbers as its universe, and interprets the function symbols as corresponding mathematical operations (i.e., multiplication and division).

As can be seen, the above causal model is cyclic with respect to its endogenous parameters P and V. However, if we fix the volume V, we obtain a submodel

$$P = c \cdot \frac{T}{V} \qquad V = v$$

This submodel provides a description of *Gay-Lussac's Law*, according to which pressure is proportional to temperature. Note, however, that this submodel is still "structural" in that the temperature of the gas is not *determined* by its pressure.

Similarly, by fixing pressure, we obtain another submodel

$$P = p \qquad V = c \cdot \frac{T}{P}$$

that represents *Charles's Law* by which the volume is proportional to temperature (though not the other way round).

3 Review: Causal Calculus

Now we will turn to a logical theory of causal reasoning that has emerged in AI — the causal calculus.

Based on the ideas from (Geffner 1992), the causal calculus was introduced in (McCain and Turner 1997) as a nonmonotonic formalism purported to serve as a logical basis for reasoning about action and change in AI. A generalization of the causal calculus to the first-order classical language was described in (Lifschitz 1997). This line of research has led to the action description language $\mathcal{C}+$, which is based on this calculus and serves for describing dynamic domains (Giunchiglia et al. 2004). A logical basis of the causal calculus was described in (Bochman 2003), while (Bochman 2004; 2007) studied its possible uses as a general-purpose nonmonotonic formalism.

Propositional case

In this section we identify a propositional interpretation ('world') with the set of propositional formulas that hold in it.

Definition 7. A *propositional causal rule* is an expression of the form $A \Rightarrow B$ (*"A causes B"*), where A and B are propositional formulas.[1] The formula A is the *body* of the rule, and B is its *head*. A *propositional causal theory* is a set of propositional causal rules.

[1](Giunchiglia et al. 2004) adopted a more cautious informal reading of such rules, namely *"If A holds, then B is caused"*.

A nonmonotonic semantics of a causal theory can be defined as follows.

Definition 8. • For a causal theory Δ and a set u of propositions, let $\Delta(u)$ denote the set of propositions that are caused by u in Δ:

$$\Delta(u) = \{B \mid A \Rightarrow B \in \Delta, \text{ for some } A \in u\}$$

• A world α is an *exact model* of a causal theory Δ if it is the unique model of $\Delta(\alpha)$. The set of exact models forms a *nonmonotonic semantics* of Δ.

The above nonmonotonic semantics of causal theories is equivalent to the semantics described in (McCain and Turner 1997). It can be verified that exact models of a causal theory are precisely the worlds that satisfy the condition

$$\alpha = \mathrm{Th}(\Delta(\alpha)),$$

where Th is the logical closure operator of classical propositional logic. Informally speaking, an exact model is a world that is closed with respect to the causal rules and also has the property that any proposition that holds in it is caused (determined) ultimately by other propositions.

Definite causal theories and completion

A propositional causal theory is *determinate* if the head of each of its rules is a literal or the falsity constant $\mathbf{f}$. A causal theory is called *definite* if it is determinate and no literal is the head of infinitely many rules of Δ. It turns out that the nonmonotonic semantics of a definite causal theory Δ coincides with the classical semantics of the propositional theory obtained from Δ by a syntactic transformation similar to program completion (Clark 1978).

The *(literal) completion* of a definite causal theory Δ is the set of all classical formulas of the forms

$$p \leftrightarrow \bigvee\{A \mid A \Rightarrow p \in \Delta\}$$
$$\neg p \leftrightarrow \bigvee\{A \mid A \Rightarrow \neg p \in \Delta\},$$

for any atom p, plus the set $\{\neg A \mid A \Rightarrow \mathbf{f} \in \Delta\}$.

As proved in (McCain and Turner 1997), the completion of a determinate causal theory provides a classical logical description of its nonmonotonic semantics:

Proposition 1. *The nonmonotonic semantics of a definite causal theory coincides with the classical semantics of its completion.*

It should be kept in mind, however, that this completion transformation is not modular with respect to the causal rules of the source theory and, moreover, it changes nonmonotonically with the changes of the latter. Speaking generally, the completion (as well as the nonmonotonic semantics itself) does not fully represent the *logical* content of a causal theory. This distinction between logical and nonmonotonic aspects of a causal theory bears immediate relevance to the distinction between causal and purely mathematical understanding of structural equations in Pearl's theory.

The logical basis of the causal calculus

The causal calculus, like other nonmonotonic formalisms, can be viewed as a two-layered construction. The nonmonotonic semantics defined above forms its top level. Its bottom level is the monotonic logic of causal rules introduced in (Bochman 2003; 2004); it constitutes the *causal logic* of the causal calculus.

In the following definition, $\models$ denotes entailment in the sense of classical propositional logic.

A *causal inference relation* is a relation $\Rightarrow$ on the set of propositions satisfying the following conditions:

(Strengthening) If $A \models B$ and $B \Rightarrow C$, then $A \Rightarrow C$;

(Weakening) If $A \Rightarrow B$ and $B \models C$, then $A \Rightarrow C$;

(And) If $A \Rightarrow B$ and $A \Rightarrow C$, then $A \Rightarrow B \wedge C$;

(Or) If $A \Rightarrow C$ and $B \Rightarrow C$, then $A \vee B \Rightarrow C$;

(Cut) If $A \Rightarrow B$ and $A \wedge B \Rightarrow C$, then $A \Rightarrow C$;

(Truth) $\mathbf{t} \Rightarrow \mathbf{t}$;

(Falsity) $\mathbf{f} \Rightarrow \mathbf{f}$.

Causal inference relations satisfy almost all the usual postulates of classical inference, except Reflexivity $A \Rightarrow A$. The absence of the latter has turned out to be essential for an adequate representation of causal reasoning.

A possible worlds semantics. A logical semantics of causal inference relations has been given in (Bochman 2004) in terms of possible worlds (Kripke) models.

Definition 9. A causal rule $A \Rightarrow B$ is said to be *valid* in a Kripke model (W, R, V) if, for any worlds α, β such that $R\alpha\beta$, if A holds in α, then B holds in β.

It has been shown that causal inference relations are complete for quasi-reflexive Kripke models, that is, for Kripke models in which the accessibility relation R satisfies the condition that if $R\alpha\beta$, for some β, then $R\alpha\alpha$.

The above semantics sanctions a simple modal representation of causal rules. Namely, the validity of $A \Rightarrow B$ in a possible worlds model is equivalent to validity of the formula $A \rightarrow \Box B$, where $\Box$ is the standard modal operator. In fact, this modal representation has been used in many other approaches to formalizing causation in action theories (see, e.g., (Geffner 1990; Turner 1999; Giordano, Martelli, and Schwind 2000; Zhang and Foo 2001)).

Strong equivalence. It has been shown in (Bochman 2003) that if $\Rightarrow_\Delta$ is the least causal inference relation that includes a causal theory Δ, then $\Rightarrow_\Delta$ has the same nonmonotonic semantics as Δ. This has shown that the rules of causal inference are adequate for reasoning with respect to the nonmonotonic semantics. Moreover, as a consequence of a corresponding *strong equivalence* theorem, it was shown that the above causal inference relations constitute a maximal such logic.

Definition 10. Causal theories Γ and Δ are called

- *objectively equivalent* if they have the same nonmonotonic semantics;
- *strongly equivalent* if, for any set Φ of causal rules, $\Delta \cup \Phi$ is objectively equivalent to $\Gamma \cup \Phi$;
- *causally equivalent* if $\Rightarrow_\Delta = \Rightarrow_\Gamma$.

Two causal theories are causally equivalent if each theory can be obtained from the other using the inference postulates of causal relations. Then the following result has been proved in (Bochman 2004):

Proposition 2 (Strong equivalence)**.** *Causal theories are strongly equivalent iff they are causally equivalent.*

First-order case

According to (Lifschitz 1997), a *first-order causal rule* is an expression of the form $G \Rightarrow F$, where F and G are first-order formulas. A *first-order causal theory* Δ is a finite set of first-order causal rules coupled with a list $\mathbf{c}$ of object, function and/or predicate constants, called the *explainable* symbols of Δ.

The nonmonotonic semantics of first-order causal theories was defined in (Lifschitz 1997) by a syntactic transformation that turns Δ into a second-order sentence D_Δ. That sentence provides a precise formal description of the requirement that the explainable symbols should be explained, or *determined*, by Δ.

This transformation is defined as follows. Let $v\mathbf{c}$ denote a list of new variables similar to $\mathbf{c}$,[2] and let $\Delta(v\mathbf{c})$ denote the conjunction of the formulas

$$\forall \mathbf{x}(G \rightarrow F^{\mathbf{c}}_{v\mathbf{c}})$$

for all rules $G \Rightarrow F$ of Δ, where $\mathbf{x}$ is the list of all free variables of F, G, and $F^{\mathbf{c}}_{v\mathbf{c}}$ denotes the result of substituting the variables $v\mathbf{c}$ for the corresponding constants $\mathbf{c}$ in F. Then D_Δ is the second-order sentence

$$\forall v\mathbf{c}(\Delta(v\mathbf{c}) \leftrightarrow (v\mathbf{c} = \mathbf{c})).$$

The sentence D_Δ (and its classical interpretations) is viewed then as describing the nonmonotonic semantics of the causal theory Δ. Informally speaking, these are the models of Δ in which the interpretation of the explainable symbols $\mathbf{c}$ is the only interpretation of these symbols that is determined, or "causally explained," by the rules of Δ.

The process of literal completion, defined for definite propositional causal theories, is extended to two classes of first-order causal theories in (Lifschitz 1997) and (Lifschitz and Yang 2013). We consider here the special case of the definition from the second paper when every explainable symbol of Δ is an object constant, and Δ consists of rules of the form

$$G(x) \Rightarrow c = x,$$

one for each explainable symbol c, where $G(x)$ is a formula without any free variables other than x. The *(functional) completion* of Δ is defined in this case as the conjunction of the sentences

$$\forall x(c = x \leftrightarrow G(x))$$

[2] That is to say, the lists $\mathbf{c}$ and $v\mathbf{c}$ have the same length; object constants in the former correspond to object variables in the latter, function symbols correspond to function variables, and predicate symbols to predicate variables.

for all rules of Δ.

The first-order causal calculus is closer to the causal models of Pearl than the propositional causal calculus, not only because it is based on a richer language, but also because it relaxes the requirement of total explainability of the latter, and restricts it to explainable symbols only. It has been noted in (Lifschitz 1997), however, that we can easily turn, for example, an unexplainable (exogenous) predicate $Q(x)$ into an explainable predicate by adding the following two causal rules:

$$Q(x) \Rightarrow Q(x) \qquad \neg Q(x) \Rightarrow \neg Q(x).$$

This will not change the nonmonotonic semantics. Still, it will allow us to reduce partial explainability to the universal explainability of the propositional causal calculus.

Describing the monotonic (logical) basis of the first-order causal calculus remains at this point an open problem.

4 Representing Structural Equations by Causal Rules

We will describe now a formal representation of Pearl's causal models as causal theories[3]. The representation itself is fully modular, and the nonmonotonic semantics of the resulting causal theory corresponds to the solutions of the causal model.

Propositional case

Definition 11. For any Boolean causal model M, Δ_M is the propositional causal theory consisting of the rules

$$F \Rightarrow A \text{ and } \neg F \Rightarrow \neg A$$

for all equations $A = F$ in M and the rules

$$A \Rightarrow A \text{ and } \neg A \Rightarrow \neg A$$

for all exogenous atoms A of M.

Theorem 3. *The causal worlds of a Boolean causal model M are identical to the exact models of Δ_M.*

Remark. The above representation was chosen from a number of alternative (logically non-equivalent) translations producing the same nonmonotonic semantics. The choice reflected Pearl's dictum that both truth and falsity assignments to an endogenous atom should be determined by the corresponding function. It has turned out to be adequate also for establishing a logical correspondence between the two formalisms (described in the next section).

Example 1, continued. If M is the causal model from the firing squad example then Δ_M consists of the causal rules

$$U \Rightarrow C,\ \neg U \Rightarrow \neg C,\ C \Rightarrow A,\ \neg C \Rightarrow \neg A,$$
$$C \Rightarrow B,\ \neg C \Rightarrow \neg B,\ A \vee B \Rightarrow D,\ \neg(A \vee B) \Rightarrow \neg D,$$
$$U \Rightarrow U,\ \neg U \Rightarrow \neg U.$$

This causal theory has two exact models, identical to the solutions (causal worlds) of M.

[3]It should be noted here that a connection between structural equation models and the causal theory by McCain and Turner has been pointed out already in (Geffner 1997).

Definition 12. Given a determinate causal theory Δ, a set X of atoms, and a truth-valued function I on X, the *subtheory* Δ^I_X of Δ is the determinate causal theory obtained from Δ by (i) removing all rules $A \Rightarrow B$ and $A \Rightarrow \neg B$ with $B \in X$, (ii) adding the rule $\mathbf{t} \Rightarrow B$ for each $B \in X$ such that $I(B) = \mathbf{t}$, and (iii) adding the rule $\mathbf{t} \Rightarrow \neg B$ for each $B \in X$ such that $I(B) = \mathbf{f}$.

Subtheories of propositional causal theories exactly correspond to submodels of Boolean causal models: the causal theory $\Delta_{M^I_X}$ is essentially identical to the subtheory $(\Delta_M)^I_X$ of Δ_M. The only difference is that the former contains additional trivial rules with the body $\mathbf{f}$.

Example 1, continued. The submodel $M^I_{\{A\}}$ of M with $I(A) = \mathbf{t}$ that was used for evaluating the action sentence S4 corresponds to the subtheory $\Delta^I_{\{A\}}$:

$$U \Rightarrow C,\ \neg U \Rightarrow \neg C,\ \mathbf{t} \Rightarrow A,$$
$$C \Rightarrow B,\ \neg C \Rightarrow \neg B,\ A \vee B \Rightarrow D,\ \neg(A \vee B) \Rightarrow \neg D,$$
$$U \Rightarrow U,\ \neg U \Rightarrow \neg U.$$

First-order case

We will generalize now the above representation to a first-order language.

Definition 13. For any first-order causal model M, Δ_M is the first-order causal theory whose explainable constants are the endogenous symbols of M, and whose rules are

$$x = t \ \Rightarrow\ x = c,$$

for every structural equation $c = t$ from M (where x is a variable).

The following theorem is a key result of this study:

Theorem 4. *An extension of the interpretation of rigid and function symbols in M to the exogenous and endogenous symbols on a universe of cardinality > 1 is a solution of M iff it is a nonmonotonic model of Δ_M.*

The proof of the above result follows from the results on functional completion, described in (Lifschitz and Yang 2013).

5 Logical and Causal Correspondences

It has been shown above that Pearl's causal models are representable as causal theories of the causal calculus in such a way that the nonmonotonic semantics of the resulting causal theory corresponds to the solutions of the source structural equations. However, in order to establish a full correspondence between Pearl's causal formalism and its representation in the causal calculus, we have to show also that the causal logic associated with the causal calculus provides an adequate basis for *causal reasoning* in Pearl's theory. At this point, some more specific features and constraints of Pearl's causal models will turn out to be crucial for establishing a proper correspondence.

Our representation of causal models produces quite specific causal theories. More precisely, in the propositional case it implies, in effect, that for any explainable atom p there exists a propositional formula F in which p does not

occur, such that $F \Rightarrow p$ and $\neg F \Rightarrow \neg p$ are the only causal rules of the corresponding causal theory that involve p in heads. A slight generalization of these restrictions will lead us to the following special kind of causal theories:

Definition 14. A propositional causal theory will be called a *causal Pearl theory* if it is determinate and satisfies the following conditions:

- no atom can appear both in the head and the body of a causal rule;
- two rules $A \Rightarrow p$ and $B \Rightarrow \neg p$ belong to a causal theory only if $A \wedge B$ is classically inconsistent.

The above class of causal theories will be sufficient for the correspondence results, given below. It should be noted, however, that the second condition above can be traced back at least to (Darwiche and Pearl 1994), where it played an essential role in constructing symbolic causal networks satisfying the Markov property.

Manipulability vs. causal equivalence

On Pearl's view, the *causal* content of a causal model is revealed in forming its submodels. We will try to formalize this understanding by introducing the following

Definition 15. Determinate causal theories Γ and Δ will be called *intervention-equivalent* if, for every set X of atoms and every truth-valued function I, the subtheory Γ^I_X has the same nonmonotonic semantics as the subtheory Δ^I_X.

On our 'reconstruction' of Pearl's views, intervention-equivalent causal theories must have the same causal content, since they determine the same answers for all possible interventions. Given this understanding, it is only natural to inquire how this notion of equivalence is related to the logical notion of causal equivalence from the causal calculus. The comparison is not straightforward, however, since causal equivalence characterizes the behavior of causal theories with respect to *expansions* of a causal theory with new rules (see Proposition 2). In contrast, causal Pearl theories do not easily admit additions of new rules at all. Despite this, the following key result gives us one direction of the correspondence:

Theorem 5. *Causal Pearl theories are intervention-equivalent only if they are causally equivalent.*

It should be noted, however, that the above result ceases to hold for more general causal theories:

Example 3. Causal theories $\{p \Rightarrow p\}$ and $\{\mathbf{t} \Rightarrow p\}$ are not causally equivalent. Still, it is easy to verify that they, and all their subtheories, have the same nonmonotonic semantics.

Unfortunately, causal equivalence does not imply intervention-equivalence, and the reason is that causal equivalence is in general not preserved by subtheories. The following example is instructive in showing why this happens:

Example 4. Let us consider two causal theories $\{p \Rightarrow \neg q,\ \ r \Rightarrow s\}$ and $\{p \Rightarrow \neg q,\ \ r \wedge \neg(p \wedge q) \Rightarrow s\}$. It is easy to show that these theories are causally equivalent. However, if we fix q, we obtain non-equivalent subtheories $\{\mathbf{t} \Rightarrow q, r \Rightarrow s\}$ and $\{\mathbf{t} \Rightarrow q, r \wedge \neg(p \wedge q) \Rightarrow s\}$.

The intervention produced non-equivalent theories because the antecedent of the rule $r \wedge \neg(p \wedge q) \Rightarrow s$ contained a superfluous part $\neg(p \wedge q)$ that could be eliminated using the first rule $p \Rightarrow \neg q$, but has become non-superfluous when the latter rule has been removed. Speaking generally, Pearl's interventions are sensitive to the presence of redundant parameters in rules (as well as in structural equations).

Fortunately, a so far neglected further restriction appearing in Pearl's description of a causal model turns out to be essential for providing a proper correspondence; it is the requirement that every structural equation should involve only a *minimal* set of parameters (= parents) sufficient for representing the corresponding function f_i. It should be kept in mind, however, that the process of finding such a minimal set of parameters is essentially derivational, because it involves exploring possible substitutions of endogenous parameters in equations by their determining functions. Below we will introduce a counterpart of this restriction for causal theories.

Definition 16.
- A rule $A \Rightarrow l$ of a causal theory Δ will be called *modular in* Δ if $B \models A$ whenever $B \Rightarrow_\Delta l$.
- A determinate causal theory will be called *modular* if all its rules are modular.

A causal rule $A \Rightarrow l$ is modular in a causal theory Δ if its body A constitutes a logically *weakest cause* of the literal l with respect to Δ[4]. Then we obtain

Theorem 6. *Causally equivalent modular causal theories are intervention-equivalent.*

The above two theorems jointly lead to the following

Corollary 7. *Modular causal Pearl theories are intervention-equivalent iff they are causally equivalent.*

The above result establishes, in effect, a required correspondence between Pearl's manipulative account of causation and its logical counterpart in the causal calculus.

6 Summary and Prospects

It was noted by Pearl in the Epilogue to (Pearl 2000), 'The Art and Science of Cause and Effect', that many scientific discoveries have been delayed for the lack of a mathematical language that can amplify ideas and let scientists communicate results. In this respect, we hope that our representation has provided a missing *logical* language for Pearl's causal reasoning, a language that would return this reasoning back to logic, albeit a nonmonotonic one.

The fact that Pearl's causal models are interpreted in this study as *nonmonotonic* causal theories allows us to clarify a large part of confusions surrounding a causal reading of structural equations. As we discussed in Section 3, the causal calculus is a two-layered construction. On the first level it has the underlying *causal logic*, a fully logical, though non-classical, formalism of causal inference relations that has its own (possible worlds) *logical semantics*. This causal logic and its semantics provide a formal interpretation for the

[4] A 'brute' way of constructing such a modular rule for a (finite) causal theory could consist in finding all minimal conjunctions of literals that cause l and forming their disjunction.

causal rules. Above this layer, however, the causal calculus includes a nonmonotonic 'overhead' that is determined by the corresponding *nonmonotonic semantics*. Furthermore, in our particular case, this nonmonotonic semantics can be captured by classical logical means, namely by forming the *completion* of the source causal theory, which is an ordinary classical logical theory that provides a complete description for the nonmonotonic semantics. In the first-order case it directly describes the corresponding equations in the usual mathematical sense. Still, this completion construction is global and non-modular with respect to the source causal theory, and it changes nonmonotonically with changes of the latter. That is why the nonmonotonic causal reasoning is not reducible to a standard logical reasoning[5].

Despite the established connection between the two causal formalisms, there are obvious differences in the respective objectives of these theories, as well as in required expressive means. Thus, the restrictions appearing in our definition of a causal Pearl theory make such theories completely inadequate for describing dynamic domains in reasoning about action and change. Consequently, an 'attuning' of the causal calculus to the demands of Pearl's theory would obviously require a significant effort in developing the necessary formal tools and appropriate reasoning mechanisms. Nevertheless, the suggested representation constitutes a highly expressive logical replacement of structural equations for many of the purposes envisaged by Pearl. We are planning to explore this representation for studying the key concepts of Pearl's theory, such as interventions (actions), counterfactuals, actual causation and explanations. For instance, the logical generality offered by the causal calculus could be exploited in analyzing Pearl's approach to counterfactuals, without restrictions to recursive (acyclic) causal models or unique solutions (cf. (Halpern 2000)), and even for extending it to counterfactuals with arbitrary antecedents, in contrast to Pearl's representation that restricts them to conjunctions of atoms, thus preventing an analysis of disjunctive counterfactuals such as "If Bizet and Verdi were compatriots...".

Acknowledgements

Thanks to Amelia Harrison for comments on the draft of this paper. The second author was partially supported by the National Science Foundation under Grant IIS-1422455.

References

Bochman, A. 2003. A logic for causal reasoning. In *IJCAI-03, Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, Mexico, August 9-15, 2003*, 141–146. Acapulco: Morgan Kaufmann.

Bochman, A. 2004. A causal approach to nonmonotonic reasoning. *Artificial Intelligence* 160:105–143.

Bochman, A. 2007. A causal theory of abduction. *Journal of Logic and Computation* 17:851–869.

[5]See (Bochman 2011) for a general discussion of the role of logic in nonmonotonic formalisms.

Bochman, A. 2011. Logic in nonmonotonic reasoning. In Brewka, G.; Marek, V. W.; and Truszczynski, M., eds., *Nonmonotonic Reasoning. Essays Celebrating its 30th Anniversary*. College Publ. 25–61.

Clark, K. 1978. Negation as failure. In Gallaire, H., and Minker, J., eds., *Logic and Data Bases*. Plenum Press. 293–322.

Darwiche, A., and Pearl, J. 1994. Symbolic causal networks. In *Proceedings AAAI'94*, 238–244.

Geffner, H. 1990. Causal theories for nonmonotonic reasoning. In *Proceedings of National Conference on Artificial Intelligence (AAAI)*, 524–530. AAAI Press.

Geffner, H. 1992. *Default Reasoning. Causal and Conditional Theories*. MIT Press.

Geffner, H. 1997. Causality, constraints and the indirect effects of actions. In *Proceedings Int. Joint Conf. on Artificial Intelligence, IJCAI'97*, 555–561.

Giordano, L.; Martelli, A.; and Schwind, C. 2000. Ramification and causality in a modal action logic. *J. Log. Comput.* 10(5):625–662.

Giunchiglia, E.; Lee, J.; Lifschitz, V.; McCain, N.; and Turner, H. 2004. Nonmonotonic causal theories. *Artificial Intelligence* 153:49–104.

Halpern, J. Y. 2000. Axiomatizing causal reasoning. *Journal of AI Research* 12:317–337.

Lifschitz, V., and Yang, F. 2013. Functional completion. *Journal of Applied Non-Classical Logics* 23(1-2):121–130.

Lifschitz, V. 1997. On the logic of causal explanation. *Artificial Intelligence* 96:451–465.

McCain, N., and Turner, H. 1997. Causal theories of action and change. In *Proceedings AAAI-97*, 460–465.

Pearl, J. 2000. *Causality: Models, Reasoning and Inference*. Cambridge UP. 2nd ed., 2009.

Pearl, J. 2012. The causal foundations of structural equation modeling. In Hoyle, R. H., ed., *Handbook of Structural Equation Modeling*. New York: Guilford Press. chapter 5, 68–91.

Turner, H. 1999. A logic of universal causation. *Artificial Intelligence* 113:87–123.

Zhang, D., and Foo, N. 2001. EPDL: A logic for causal reasoning. In *Proc Int. Joint Conf. on Artificial Intelligence, IJCAI-01*, 131–136. Seattle: Morgan Kaufmann.