Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Default Logic as a Species of Causal Reasoning

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

Abstract

We will show that Reiter's default logic can be viewed as a particular instantiation of causal reasoning. This will be demonstrated by establishing back and forth translations between default theories and causal theories of the causal calculus, using a particular causal nonmonotonic semantics called a default semantics. Moreover, it will be shown that Pearl's structural equation models can be viewed as default causal theories in this sense. We will discuss also some global consequences this representation could have for establishing a general role of causation in nonmonotonic reasoning.

1 Introduction

The field of nonmonotonic reasoning has undergone an extensive development in its relatively short, less than a half-century life. However, this impressive development has also brought about an unbridled proliferation of formalisms and approaches to nonmonotonic reasoning. Already now, there are many areas of what is still customarily called nonmonotonic reasoning that do not, and even cannot communicate with each other. In this respect, the present purely theoretical study can be viewed as a contribution to an opposite, unification development of this field. More precisely, our ultimate aim consists in showing that significant parts of nonmonotonic reasoning can be viewed as different instantiations of general causal reasoning. This study will contribute to this task by demonstrating that one of the key, original formalisms of nonmonotonic reasoning, default logic of Raymond Reiter (**?**), can be viewed as such an instantiation of causal reasoning. The corresponding causal representation will also allow us to clarify the meaning of the main notions associated with default logic and first of all of the concept of default itself.

Our basic language in this paper will be a classical propositional language with the usual classical connectives and constants $\{\wedge, \vee, \neg, \to, \mathbf{t}, \mathbf{f}\}$. The symbol $\vDash$ will stand for the classical entailment while Th will denote the associated classical provability operator. In this study, $p, g, r, \dots$ will usually denote propositional atoms while $A, B, C, \dots$ will denote arbitrary classical propositions.

2 Default Logic

Default logic has been born as just one of a number of alternative formalisms for nonmonotonic reasoning. Still, it seems safe to argue even today that it is one of the key formalisms in this broad area of reasoning, both with respect to its representation capabilities and in its relations with other nonmonotonic formalisms (such as logic programming).

Originally, a default theory was defined in (**?**) as a pair (W, D), where W is a set of classical propositions (the axioms) and D a set of *default rules* having an unusual form $A : b/C$, where A and C are propositions and b a finite set of propositions (called justifications). Very informally, a rule $A : b/C$ was intended to state something like:

> "If A is accepted, and each $B \in b$ can be consistently assumed, then C should be accepted."

The most salient feature of such rules was their *defeasibility*; a default rule was viewed as an inference rule $A \vdash C$ that should always be applied unless it is canceled by refuting one of its justifications in b.

Default rules were intended to act as meta-rules for extending the initial knowledge base W beyond what is strictly known. Accordingly, the nonmonotonic semantics of default logic was defined by determining admissible *extensions* of a default theory. An extension was defined by a fixed point construction: for a set u of propositions, let $\Gamma(u)$ be the least deductively closed set that includes W and satisfies the following condition:

> If $A : b/C \in D$, $A \in \Gamma(u)$ and $\neg B \notin u$, for every $B \in b$, then $C \in \Gamma(u)$.

Then a set s is an *extension* of the default theory if and only if $\Gamma(s) = s$.

(**?**) suggested a more abstract description of default logic using the notion of a context-dependent proof as a way of formalizing Reiter's operator Γ. This representation was developed in (**?**) to a general theory of nonmonotonic rule systems (see also (**?**)).

Given a set s of propositions (the "context"), let us consider the set $\mathcal{D}(s)$ of all propositions that are derivable from W using classical entailment and the following ordinary inference rules that are allowed by the context:

$$\{A \vdash C \mid A : b/C \in D \;\&\; \neg B \notin s, \text{ for every } B \in b\}.$$

Then s is an extension of the default theory if and only if $s = \mathcal{D}(s)$.

The above representation has made it vivid that a large part of reasoning in default logic involves ordinary rule-based deductive inference, the only distinction from traditional deductive systems being that the very set of rules allowed in the inference process is determined by the (assumptions made in the) context. In particular, an extension of a default theory can be viewed as a set of propositions that are logically provable (= justified) on the basis of taking itself as an assumption context.

In what follows, instead of a pair (W, D) of propositional axioms and default inference rules, we will represent a default theory simply as a set of rules of the form $A : b/C$, where b may be empty. In the latter case, the default rule $A : /C$ with an empty justification set represents an ordinary inference rule $A \vdash C$, while the axioms W are representable via rules of the form $:/A$, for $A \in W$.

2.1 Default Logic Made Simple

It was shown in (**?**) that default logic is reducible to a rather simple formalism that contains only deductive inference rules and default assumptions.

The reduction. Let us extend the source classical propositional language $\mathcal{L}$ with new propositional atoms A° for every proposition A in $\mathcal{L}$. For a set u of propositions from $\mathcal{L}$, u° will denote the set of new atoms $\{A^\circ \mid A \in u\}$.

Next, if D is a default theory in $\mathcal{L}$, then D° will denote a default theory in the extended language that includes the following set of plain inference rules:

$$\{A, b^\circ \vdash C \mid A : b/C \in D\},$$

plus the following two rules for every formula A from $\mathcal{L}$ that appears as a justification in the rules from D:

$$\neg A \vdash \neg A^\circ \qquad \text{and} \qquad : A^\circ/A^\circ.$$

The following theorem (proved in (**?**)) shows, in effect, that the above translation is a polynomial, faithful, and modular translation (PFM) in the sense of (**?**).

Proposition 1. *A set u is an extension of a default theory D if and only if there is a unique extension u_0 of D° such that $u = u_0 \cap \mathcal{L}$.*

The above translation reduces an arbitrary default theory to a default theory containing only plain deductive inference rules and *supernormal* default rules of the form $:A/A$. We will call such default theories *simple* in what follows. The corresponding reduction allows us to provide an alternative description of default reasoning as deductive reasoning in the presence of default assumptions.

Simple default theories. Supernormal default rule $:A/A$ asserts, in effect, that proposition A should be accepted whenever it is consistent with the rest of the facts and rules of a default theory. Such a proposition can be viewed as a *default assumption* of the theory, which allows us to reformulate any simple default theory as a formalism that contain only plain inference rules and defaults.

Definition 1. A *simple default theory* is a pair $(\Delta, \mathcal{D})$, where Δ is a set of inference rules, and $\mathcal{D}$ a distinguished set of propositions called *default assumptions* (or simply *defaults*).

Default reasoning in this setting amounts to deriving justified conclusions from a default theory by using its rules and default assumptions. However, in the case when the set of all defaults $\mathcal{D}$ is jointly incompatible with the background theory Δ, we must make a reasoned choice among the default assumptions. At this point, default reasoning requires that a reasonable set of assumptions that can be actually used in this context not only should be consistent and maximal but also should explain why the rest of the default assumptions should be rejected. An important prerequisite of such explanations is that the underlying inference system Δ contains *cancellation* rules by which some sets of assumptions refute others (given the known facts). The appropriate choices of assumptions will determine then the nonmonotonic semantics of a default theory.

For an arbitrary set Δ of deductive, Tarski inference rules, let Cn_Δ denote the provability operator associated with the least supraclassical consequence relation containing Δ. In other words, for any set u of propositions, $\mathrm{Cn}_\Delta(u)$ is the set of propositions that are derivable from u using the rules from Δ and classical entailment.

Definition 2. Given a simple default theory $(\Delta, \mathcal{D})$,

- A set $\mathcal{D}_0$ of defaults will be called *stable* in $(\Delta, \mathcal{D})$ if it is consistent and refutes any default outside the set:

$$\neg A \in \mathrm{Cn}_\Delta(\mathcal{D}_0), \text{ for any } A \in \mathcal{D} \setminus \mathcal{D}_0.$$

- A set s of propositions is an *extension* of a simple default theory if $s = \mathrm{Cn}_\Delta(\mathcal{D}_0)$ for some stable set of defaults $\mathcal{D}_0$. The set of extensions determines the *nonmonotonic semantics* of the default theory.

Combining the above definitions of a stable set and that of extension, we obtain the following description of the nonmonotonic semantics.

Proposition 2. *A set s of propositions is an extension of a simple default theory $(\Delta, \mathcal{D})$ if and only if it satisfies the following two conditions:*

- *s is the deductive closure of the set of its defaults:*

$$s = \mathrm{Cn}_\Delta(\mathcal{D} \cap s);$$

- *s decides the default set: for any $A \in \mathcal{D}$, either $A \in s$, or $\neg A \in s$.*

Simple default theories and their nonmonotonic semantics provide presumably the simplest and most transparent description of a full-fledged nonmonotonic reasoning (that is, of nonmonotonic reasoning that includes classical deductive logic as its essential logical basis). In a broader perspective, however, the above description also displays default logic as a particular instantiation of two more general approaches to our reasoning.

To begin with, simple default theories display default logic as a particular instance of an *abductive reasoning* in which default assumptions play the role of abducibles. This understanding of default logic was envisaged to some extent already in the Theorist system of (?). Still, the abductive formalism of Theorist has not fully captured the representation capabilities of default logic due to the fact that it was fully classical, while default reasoning is only supraclassical; though it subsumes classical logic, it also requires essential use of inference rules that are not reducible to corresponding classical implications.

As a second correspondence, the above formalism makes it especially vivid that default logic can be viewed as a principal instantiation of a formal *assumption-based argumentation* (see (?)) where default assumptions play the role of arguments. We will return to these connections with general reasoning formalisms repeatedly in what follows.

3 The Causal Calculus

Based on ideas from (?), the causal calculus was introduced in (?) 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 (?). 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 (?). A logical basis of the causal calculus was described in (?) while (?; ?) studied its possible uses as a general-purpose nonmonotonic formalism. Later it was shown in (?) that structural equation models of (?) are representable in the causal calculus, so the latter can actually be seen as a formalism that provides a unified logical approach to causation. A more systematic description of this approach, its historical roots, and the range of its possible applications in nonmonotonic reasoning and beyond can be found in (?).

The causal calculus can be seen as a natural extension of classical logic that allows for causal reasoning. From a purely logical point of view, this generalization amounts to dropping the reflexivity postulate of classical deductive inference. However, the associated inference systems are assigned both an 'ordinary' *logical* semantics (that gives a semantic interpretation to causal rules) and a natural *nonmonotonic* semantics which provides a representation framework for causal reasoning.

3.1 Production Inference

The basic informational units of the causal calculus are rules, or conditionals, of the form $A \Rightarrow B$ that hold among classical propositions. A rule $A \Rightarrow B$ says that A *causes* B. Such rules determine our causal language, which is built on top of the underlying language of classical logic. Moreover, the main role of the postulates in the definition below amounts to securing that the corresponding causal reasoning respects this underlying classical logic of propositions.

Definition 3. A *production inference relation* is a binary relation $\Rightarrow$ on the set of classical propositions satisfying the following postulates:

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

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

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

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

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

The most significant "omission" of the above set of postulates is the absence of the reflexivity postulate $A \Rightarrow A$. It is precisely this feature of causal rules that creates the possibility of nonmonotonic reasoning in this framework.

We extend causal rules to rules having arbitrary sets of propositions as premises using the compactness recipe: for any set u of propositions, we define $u \Rightarrow A$ as follows:

$$u \Rightarrow A \;\equiv\; \bigwedge a \Rightarrow A, \text{ for some finite } a \subseteq u.$$

For a set u of propositions, $\mathcal{C}(u)$ denotes the set of propositions caused by u, that is,

$$\mathcal{C}(u) = \{A \mid u \Rightarrow A\}.$$

The causal operator $\mathcal{C}$ plays much the same role as the usual derivability operator for consequence relations. Note, in particular, that it is monotonic.

(Monotonicity) If $u \subseteq v$, then $\mathcal{C}(u) \subseteq \mathcal{C}(v)$.

Note also that $\mathcal{C}(u)$ is always a deductively closed set: for any set u,

$$\mathcal{C}(u) = \mathrm{Th}(\mathcal{C}(u)).$$

Still, $\mathcal{C}$ is not inclusive: that is, $u \subseteq \mathcal{C}(u)$ does not always hold. Also, it is not idempotent: that is, $\mathcal{C}(\mathcal{C}(u))$ can be distinct from $\mathcal{C}(u)$.

Causal theories. By a *causal theory* we will mean an arbitrary set of causal rules.

For any set u of propositions and a causal theory Δ, we will denote by $\Delta(u)$ the set of all propositions that are directly caused by u in Δ, that is,

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

For any causal theory Δ, there exists a least production relation that includes Δ. We will denote it by $\Rightarrow_\Delta$ while $\mathcal{C}_\Delta$ will denote the corresponding causal derivability operator. Clearly, $\Rightarrow_\Delta$ is the set of all causal rules that can be derived from Δ using the postulates for a production inference relation.

Logical Semantics of Causal Rules. The semantic framework for production relations can be built from pairs of deductively closed theories called bimodels.

Definition 4. A pair of consistent deductively closed sets will be called a *classical bimodel*. A set of classical bimodels will be called a *classical binary semantics*.

A classical binary semantics can also be viewed as a *binary relation* on the set of deductive theories. Accordingly, given a set of bimodels (semantics) $\mathcal{B}$, we will write $u\mathcal{B}v$ to denote the fact that the a bimodel (u, v) belongs to $\mathcal{B}$. These descriptions will be used interchangeably in what follows.

Now we will define the notion of validity of causal rules with respect to a classical binary semantics.

Definition 5. A rule $A \Rightarrow B$ will be said to be *valid* in a classical binary semantics $\mathcal{B}$ if, for any bimodel (u, v) from $\mathcal{B}$, $A \in v$ only if $B \in u$.

It can be shown that a binary relation $\Rightarrow$ on the set of propositions is a production inference relation if and only if it is determined by a classical binary semantics.

3.2 Regular Inference

A production inference relation will be called *regular* if it satisfies:

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

Cut is one of the basic rules for ordinary consequence relations. It corresponds to the following characteristic property of the causal operator:

$$\mathcal{C}(u \cup \mathcal{C}(u)) \subseteq \mathcal{C}(u).$$

The semantic characterization of regular inference relations can be obtained by considering only classical bimodels (u, v) such that $u \subseteq v$. Such bimodels (and the corresponding semantics) have been called *consistent* ones.

Regular production relations have a number of additional properties. Thus, any such relation will already be transitive.

A causal rule of the form $A \Rightarrow \mathbf{f}$ is called a *constraint*. Such rules can be used for incorporating a purely factual information into causal theories: a rule $A \Rightarrow \mathbf{f}$ says, in a sense, that A should not hold in any intended model.

Now, an important property of regular production relations is that any causal rule implies the corresponding constraint:

(Constraint) If $A \Rightarrow B$, then $A \wedge \neg B \Rightarrow \mathbf{f}$.

Regular inference relations can already be described in terms of a usual notion of a propositional theory.

Definition 6. A set u of propositions will be called a *theory* of a production relation if it is deductively closed and $\mathcal{C}(u) \subseteq u$. A set u will be called a (propositional) *theory* of a causal theory Δ if it is a theory of $\Rightarrow_\Delta$.

A theory of a production relation is a set of propositions that is closed both deductively and with respect to its causal rules: namely, if $A \in u$ and $A \Rightarrow B$, then $B \in u$. Accordingly, such theories have much the same properties as ordinary theories of consequence relations. Note, in particular, that the set of theories of a production inference relation is closed with respect to arbitrary intersections, and consequently any set of propositions is included in the least such theory.

As could be expected, propositional theories of a causal theory Δ are sets of propositions that are closed with respect to the rules of Δ.

3.3 Causation versus Consequence

A further insight into the properties of regular causal inference can be obtained by comparing it with associated consequence relations.

Note that any causal theory, and hence any production inference relation, can also be considered as an ordinary conditional theory (a set of inference rules), so it determines the corresponding supraclassical consequence relation. In order to construct such a consequence relation, we need only to "restore" reflexivity of the corresponding inference. The following construction gives a direct description of this consequence relation in terms of the source production relation. Namely, for a (regular) production relation $\Rightarrow$, we can define the following *consequence* relation:

$$A \vdash_{\Rightarrow} B \;\equiv\; A \Rightarrow (A \rightarrow B).$$

Proposition 3. *If $\Rightarrow$ is a regular production relation, then $\vdash_{\Rightarrow}$ is the least supraclassical consequence relation containing $\Rightarrow$.*

Let $\mathrm{Cn}_{\Rightarrow}$ denote the provability operator corresponding to $\vdash_{\Rightarrow}$. Then the above description can be extended to the following equality, for any set u of propositions:

$$\mathrm{Cn}_{\Rightarrow}(u) = \mathrm{Th}(u \cup \mathcal{C}(u)).$$

Incidentally, the above equality shows that regular causal inference and its associated deductive consequence are indeed close relatives since it shows that $\mathcal{C}(u)$ captures all nontrivial consequences included in $\mathrm{Cn}_{\Rightarrow}(u)$, save for u itself. This fact will turn out to be useful in what follows.

As a first consequence of the above correspondence, we obtain:

Corollary 4. *Theories of $\Rightarrow$ coincide with the theories of $\vdash_{\Rightarrow}$.*

Since theories of $\vdash_{\Rightarrow}$ are exactly sets of propositions of the form $\mathrm{Cn}_{\Rightarrow}(u)$, the above result implies that such sets are precisely theories of $\Rightarrow$.

The above description of associated consequence relations can be immediately generalized to arbitrary causal (= conditional) theories. Thus, if Cn_Δ denotes the least supraclassical consequence relation containing a causal theory Δ while $\mathcal{C}_\Delta$ is the production operator of the least regular production relation containing Δ, then we have:

$$\mathrm{Cn}_\Delta(u) = \mathrm{Th}(u \cup \mathcal{C}_\Delta(u)).$$

3.4 Nonmonotonic Semantics of Causal Theories

As one of its primary objectives, a causal theory should determine the set of situations (or worlds) that satisfy the rules of the theory. However, this principal semantic function is realized in the causal calculus by assigning a causal theory a particular *nonmonotonic* semantics. By the intended interpretation, situations that satisfy a causal theory should not only be closed with respect to the causal rules of the theory, but they should also satisfy the law of causality, according to which any proposition that holds in a model should have a cause in this model.

Formally, the fact that the causal operator $\mathcal{C}$ is not reflexive creates an important distinction among theories of a production relation.

Definition 7. For a production inference relation $\Rightarrow$,

- A set u of propositions will be called *explanatory closed*, if $u \subseteq \mathcal{C}(u)$.
- A theory u of $\Rightarrow$ will be called *exact*, if it is consistent and explanatory closed, that is,

$$u = \mathcal{C}(u).$$

- A set u is an *exact theory of a causal theory* Δ, if it is an exact theory of $\Rightarrow_\Delta$.

In what follows, an exact theory will also be called an *exact model* of the corresponding causal theory or a production relation. An exact model describes a situation that not only satisfies the relevant causal rules, but also is such that every proposition that holds in it is caused by other propositions in accordance with these rules. In this sense, it directly implements the law of causation, or Leibniz's principle of sufficient reason, as part of its definition. This leads us to the following notion of a nonmonotonic semantics:

Definition 8. A *nonmonotonic semantics* of a production inference relation or a causal theory is the set of all its exact models.

All the information that can be discerned from the nonmonotonic semantics of a causal theory can be seen as nonmonotonically implied by the latter. This includes, for instance, so-called skeptical conclusions (what holds for all exact models) or credulous conclusions (what holds in at least one exact model).

Exact theories are precisely fixed points of the causal operator $\mathcal{C}$. Since the latter operator is monotonic and continuous, exact theories (and hence the nonmonotonic semantics) always exist. Thus, the general properties of monotonic operators imply that any causal theory has the least exact model, which coincides with the set of propositions that are caused by truth $\mathtt{t}$. Also, any exact model is included in a maximal exact model.

However, exact theories are not closed with respect to arbitrary intersections. Consequently, the least exact theory containing a given set of propositions does not always exist.

It has been shown in (?) that regular inference provides an adequate and maximal logical framework for reasoning with exact models.

The following technical result (see Lemma 8.21 in (?)) will play a key role in establishing the target correspondence between extensions of a default theory and a particular kind of exact models for associated causal theories.

To simplify the notation, $\mathcal{C}$ below will denote the causal operator of a regular production relation, while Cn—the associated supraclassical consequence relation (see Proposition 3 above).

Proposition 5. *For a regular production relation, the following conditions are equivalent:*

- *u is an explanatory closed set (that is, $u \subseteq \mathcal{C}(u)$);*
- *$\mathcal{C}(u) = \mathrm{Cn}(u)$;*
- *$\mathrm{Cn}(u)$ is an exact theory.*

The above lemma says in particular that causal consequences of an explanatory closed set of propositions can be captured using the associated consequence relation Cn, *and vice versa*.

As a consequence of the above lemma, we obtain also

Corollary 6. *If u is an explanatory closed set of a regular production relation, then $\mathcal{C}(u)$ ($= \mathrm{Cn}(u)$) is an exact model.*

3.5 Axioms, Assumptions, and Abducibles

The nonmonotonic semantics of causal theories is based on the law of causality, or Leibniz's principle of sufficient reason, which requires that any proposition that belongs to an exact model should have a cause that also belongs to the model. Accordingly, justification of accepted propositions constitutes an essential part of this semantic framework (see, e.g., (?) for an abstract theory of justifications).

The law of causality inevitably leads to a fundamental problem known already in antiquity as the *Agrippan trilemma*: if you do not want to accept infinite regress of causation (or justification), you should accept either uncaused or self-caused propositions. Now, in the framework of the causal calculus, there are two kinds of propositions that can play, respectively, these two roles:

Definition 9.
- A proposition A will be called an *axiom* of a causal theory Δ if the rule $\mathbf{t} \Rightarrow A$ belongs to Δ;
- A proposition A will be called an *assumption* of a causal theory if the rule $A \Rightarrow A$ belongs to it.

In clear contrast with purely deductive reasoning, both axioms and assumptions provide reasonable end-points of the justification process in causal reasoning: axioms are grounded in the Truth ($\mathtt{t}$), while assumptions naturally correspond in this sense to self-evident propositions. It is easy to show that, even for general production relations, any axiom will also be an assumption, though not vice versa. The difference between the two can be described as follows. Every axiom *must* be accepted in any reasonable model, and hence it should belong to any exact model. In contrast, any assumption *can* be included into an exact theory when it is consistent with the latter, but it does not have to be included into it. As a result, causal theories admit in general multiple exact models, depending on the assumptions we actually accept. This functionality makes assumptions much similar to *abducibles* in an abductive system. In fact, it was shown in (?) that the causal calculus provides in this sense a uniform and syntax-independent description of abductive reasoning. Moreover, it has been shown that in many regular cases (notably, in the finite case) the correspondence between causal and abductive theories is even bidirectional in the sense that the nonmonotonic semantics of a causal theory coincides with the semantics of an associated abductive system.

4 Defaults versus Facts

It turns out, however, that even the above notion of an assumption still does not provide an adequate formal representation for the concept of default as it is (implicitly) understood in default logic. Moreover, default logic implements also a more stringent, 'puritan' notion of acceptance for the rest of propositions. Accordingly, our proclaimed aim of representing default logic as a species of causal reasoning cannot be achieved without a proper formalization of these notions in causal terms.

Remark. Some distant origins of the distinction between defaults and assumptions in general can be found in the difference between two models of diagnosis, consistency-based and abductive one (see (?)). The importance of causal reasoning for a general theory of diagnosis and, in particular, for clarifying these distinctions has been shown first in (?).

In what follows, we will consider defaults to be a special kind of assumptions. With this understanding, the difference between defaults and assumptions in general can be informally described as follows: defaults are assumptions that we *must* accept unless there are reasons to the contrary.

In order to formulate the above requirement in causal terms, let as say that a proposition A is *refuted* in a causal model if the model contains a cause for the contrary proposition $\neg A$. Then we can formulate the following (still informal) principle of Default Acceptance:

Default Acceptance *A default is an assumption that is accepted whenever it is not refuted.*

The principle of Default Acceptance might be viewed as an 'anti-Leibniz' principle since it says, in effect, that a default assumption is *not* accepted only if we have reasons for its rejection. Note, however, that the original Leibniz principle of sufficient reason should still remain to hold in exact models. In particular, a proposition $\neg A$ holds in such a model only if $\neg A$ has a cause in this model (that is, when A is refuted). Accordingly, the principle of Default Acceptance in such models boils down to the principle of Default Bivalence:

Default Bivalence *For any exact model u and any default assumption A, either $A \in u$ or $\neg A \in u$.*

The above principle of default bivalence can be considered as a characteristic property of defaults. Again, this is in clear contrast with classical logical reasoning where *all* propositions are required to satisfy bivalence. Note also that any axiom of a causal theory will also be a default on this understanding (namely a default that cannot be refuted). In this sense, defaults can be viewed as an intermediary concept between axioms and assumptions in general.

Bipolarity. Turing now to the justification status of the rest of propositions in default logic, the notion of an extension of a default theory presupposes, in effect, that any such proposition should be accepted only if it is grounded, ultimately, in the set of accepted defaults. In other words, once we choose an acceptable ("stable") set of default assumptions, the rest of acceptable propositions should be derived from this set. This pertains, in particular, even to other assumptions that could belong to a (causal) theory; any such assumption becomes unacceptable unless it is derived from accepted default assumptions.

The above stringent understanding of acceptance for defaults and the rest of propositions creates, in effect, a *bipolar system of reasoning* that divides all propositions into two classes with opposite principles of acceptance. The first class contains *factual propositions* that are viewed as unacceptable unless they are derived from (or caused by) other propositions (and ultimately by accepted defaults), while the second class contains defaults that are viewed as acceptable unless they are refuted by other propositions (and, again, ultimately by other accepted defaults). It is this understanding that also makes default logic a principal instantiation of (assumption-based) argumentation (?) where *defaults play the role of arguments* (see (?) for a more detailed description of this connection).

Remark. The above description allows us to moderate, however, the main complaint against the stable semantics of default logic and logic programming raised in (?) (and exploited as a justification for the desirability of other, more relaxed nonmonotonic semantics), namely that such a semantics is overly 'opinionated' in that it requires any argument outside a stable set of arguments to be refuted (i.e., attacked) by this stable set. On our description, this requirement (formulated as the principle of default acceptance or default bivalence) is restricted to defaults only. In a sense, it is a requirement that reflects the meaning of defaults as opposed to other propositions and even assumptions in general. On our account, a default could be *defined* as a proposition that should be accepted in the absence of reasons to the contrary.

The above principles of acceptance that are sanctioned by default logic can be formulated as additional constraints that should be imposed on acceptable exact models of causal theories, and we will provide a formal description of the corresponding models in the next section. These constraints could also be viewed, however, as a certain limitation of default logic as compared with general causal reasoning (as formalized in the causal calculus). In particular, it makes default logic less suitable for other applications of nonmonotonic reasoning in AI, such as abductive reasoning (and diagnosis), or reasoning about actions (see, e.g., (?)) that seem to require the use of assumptions that are not defaults in the sense of default logic. Still, in many regular cases these two views of causal reasoning can be reconciled for a mutual benefit of both sides. It will be shown, in particular, that even Pearl's original approach to causal reasoning in the framework of structural equation models (see (?)) can be viewed as an instantiation of a default theory.

5 Default Causal Theories

In this section we are going to formulate a 'causal counterpart' of default logic in the causal calculus.

We will begin with the following definition.

Definition 10. A *default causal theory* is a pair $(\Delta, \mathcal{D})$, where Δ is a causal theory, and $\mathcal{D}$ a distinguished subset of its assumptions, called *defaults*.

As follows from the above definition, any default A from $\mathcal{D}$ is already an assumption of the relevant causal theory, that is, a rule $A \Rightarrow A$ belongs to Δ.

Our next definition describes the intended nonmonotonic semantics of a default causal theory.

Definition 11. A *default model* of a default causal theory $(\Delta, \mathcal{D})$ is an exact model m of Δ that satisfies the following two conditions:

(Default Grounding) m is caused by the set of its defaults:

$$m = \mathcal{C}_\Delta(m \cap \mathcal{D}).$$

(Default Bivalence) For any default $D \in \mathcal{D}$, either $D \in m$ or $\neg D \in m$.

By a *default semantics* of a default causal theory we will mean the set of all its default models.

It can be verified that if an arbitrary set m of propositions satisfies the condition of Default Grounding, it will already be an exact model of the corresponding causal theory Δ, that is, $m = \mathcal{C}_\Delta(m)$ will hold. Consequently, the default semantics can be viewed as a special case of the general nonmonotonic semantics of causal theories. Still, there are two reasons why the reverse inclusion does not hold in general. First, an exact model may be generated not only by defaults, but also by other assumptions (on our causal understanding of the latter). Second, even when an exact model is caused (generated) by some set of defaults, it may still not satisfy the second condition of the above definition (the principle of default bivalence). This might happen, in particular, even when the relevant set of generating defaults is maximal in the sense that it is incompatible with every other default outside this set, but the background causal theory lacks appropriate cancellation rules that would allow us to *refute* these other defaults. As an extreme case, a default causal theory may even lack default models at all (though it always has exact models).

5.1 Correspondence with Default Logic

We will establish now a bidirectional correspondence between the above notion of a default causal theory and its default semantics and 'ordinary' default theories of Reiter's default logic with their associated semantics of extensions. In establishing this correspondence, we will make an essential use of simple default theories as an intermediate concept (see Proposition 1 above).

The first direction of the correspondence is straightforward. Given a default causal theory $\mathbb{D} = (\Delta, \mathcal{D})$, we will denote by $S(\mathbb{D})$ a simple default theory $(\Delta_c, \mathcal{D})$, where Δ_c is the set of plain deductive inference rules corresponding to the causal rules of Δ. Then we have

Theorem 7. *The default semantics of a default causal theory $\mathbb{D}$ coincides with the set of extensions of $S(\mathbb{D})$.*

The proof of the above theorem is also straightforward, given the fact that the causal inference operator $\mathcal{C}_\Delta$ used in the above definition of default causal models (see Definition 11) can be safely replaced in this particular context with the corresponding operator Cn of supraclassical deductive inference (see Proposition 5).

For the reverse direction of correspondence, from default logic to causal theories, note first that the above theorem can be reversed in the sense that any simple default theory can be directly transformed into a default causal theory by simply viewing its inference rules as causal rules. This allows us to provide a representation of default logic in the causal calculus by combining the above direct correspondence between simple default theories and causal theories with our previous reduction of arbitrary default theories to simple ones (see Proposition 1 above).

Such a representation can be described directly as follows. Given an arbitrary default theory D in some propositional language $\mathcal{L}$, we will transform it first into the corresponding simple default theory $(\Delta_D, \mathcal{D})$ in an extended language $\mathcal{L}^\circ$, where

$$\Delta_D = \{A, b^\circ \vdash C \mid A : b/C \in D\} \cup \{\neg A \vdash \neg A^\circ\},$$

whereas $\mathcal{D}$ is the set of new atoms A° for every formula A that appears as a justification in the rules from D. As a second step, we will just reformulate this simple default theory as a default causal theory $C(D)$ by considering every rule of Δ_D as a causal rule. Then we obtain the following correspondence.

Theorem 8. *A set u is an extension of a default theory D if and only if there is a unique default model u_0 of $C(D)$ such that $u = u_0 \cap \mathcal{L}$.*

5.2 Pearl's Causal Models as Default Theories

A representation of Pearl's causal models in the causal calculus was provided in (**?**). It was based, however, on a certain generalization of Pearl's notion of a causal model, a generalization that has relieved it, in effect, from some of its more specific features such as acyclicity (recursiveness) and uniqueness of solutions. As an amendment to this general representation, we are going to show now that the original Pearl's causal model can actually be viewed as an instance of a default causal theory as described in this paper.

According to (**?**, Definition 7.1.1), a causal model is a triple $M = \langle U, V, F\rangle$, where

- U is a set of *exogenous* variables,
- V is a finite set $\{V_1, V_2, \ldots, V_n\}$ of *endogenous* variables that are determined by other variables in $U \cup V$, and
- 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.

F can be represented as a set of *structural* equations

$$V_i = f_i(PA_i, U_i) \quad i = 1, \ldots, n,$$

where PA_i is the unique minimal set of variables in $V \backslash \{V_i\}$ (parents of V_i) sufficient for representing f_i, and similarly for the relevant set of exogenous variables $U_i \subseteq U$. Each such equation stands for a set of "structural" equalities

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

where v_i, pa_i and u_i are, respectively, particular realizations of V_i, PA_i and U_i. Such an equality assigns a specific value v_i to a variable V_i depending on the values of its parents and relevant exogenous variables.

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 the $=$ 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."

Since Pearl's causal models have been formulated in the language of structural equations, their comprehensive logical description could be achieved only in the first-order language. The corresponding generalization of the causal calculus to a first-order language was actually described in (**?**). Still, for our current purposes we can obviate this limitation of our (propositional) causal calculus by considering the Herbrand base of this first-order language as our propositional language in this section. This Herbrand base consists of all propositions of the form $X = x$, where X is some (exogenous or endogenous) variable while x is its particular admissible value. In other words, admissible value assignments to exogenous and endogenous variables of the structural equations will be viewed as propositional atoms of the corresponding propositional language. In particular, instantiations of exogenous and endogenous variables will be called, respectively, exogenous and endogenous atoms.

Using the above reformulation, the representation of Pearl's causal models in the causal calculus, suggested in (**?**), amounted in effect to viewing each structural equality $v_i = f_i(pa_i, u_i)$ for a particular instantiation of the relevant variables as a causal rule saying that the instantiation pa_i of the parent endogenous variables PA_i and the instantiation u_i of exogenous variables U_i *causes* the instantiation $f_i(pa_i, u_i)$ of V_i:

$$PA_i = pa_i, U_i = u_i \Rightarrow V_i = f_i(pa_i, u_i).$$

In the special case when all the relevant variables are Boolean, a Boolean structural equation $p = F$ produces in this sense two causal rules

$$F \Rightarrow p \quad \text{and} \quad \neg F \Rightarrow \neg p.$$

We refer the reader to (**?**) for a more detailed (and precise) description.

A complete representation of Pearl's causal models in the causal calculus required also a determination of the causal status of the exogenous variables. In the framework of (**?**), the corresponding exogenous atoms were taken to be the *assumptions* of the resulting causal theory. In particular, in the Boolean case these exogenous atoms were required to satisfy the rules

$$p \Rightarrow p \quad \text{and} \quad \neg p \Rightarrow \neg p.$$

As we discussed earlier, assumptions are 'self-justified' propositions in a causal theory that can be accepted whenever they are consistent with the rest, though they need not be accepted. On the face of it, this stands in a clear contrast with an understanding of the role of exogenous variables in Pearl's causal models, since instantiation of the latter provides an ultimate basis of the "causal worlds" determined by such models, so maximal coherent instantiations of all such variables *should* be accepted. Still, the way out of this discrepancy, suggested in (**?**), amounted to restricting the corresponding exact models to *worlds* in the usual logical sense. In the terminology of the present paper, however, the solution amounted to requiring Bivalence for all propositions of the corresponding language. In particular, it required all assumptions of the relevant causal theory to become *default* assumptions in the sense of the present paper.

The representation of default logic in the causal calculus, described in this paper, suggests a natural and more adequate representation of Pearl's causal models, a representation that is not based on the universal principle of bivalence. Namely, we can consider a causal model in the sense of Pearl as a *default* causal theory in which exogenous atoms play the role of defaults, while endogenous atoms correspond to factual propositions. Then the principle of default acceptance will require that any default model should decide (accept or refute) all exogenous atoms, while the first condition of the definition of a default model (namely, the condition of default grounding) will provide precise formulation for Pearl's requirement that the entire set of functions F determined by structural equations forms a mapping from U to V, which means that, once a particular instantiation of the exogenous variables $U = u$ is settled, the entire set of structural equations will have a unique solution $V(u)$. Accordingly, we can formulate the following representation result.

Theorem 9. *If $\mathbb{D}_P$ is a default causal theory corresponding to a Pearl's causal model P by the above representation, then default causal models of $\mathbb{D}_P$ correspond precisely to causal worlds of P.*

It is important to observe that, insofar as the structural equations determine a unique solution $V(u)$ for all endogenous variables and for every instantiation u of the exogenous variables U (which holds, for instance, in recursive (acyclic) models), the corresponding default causal models will still be worlds in the usual logical meaning of the term. This will happen, however, not because of the imposed bivalence for all propositions, but simply due to the fact that causal rules of the default causal theory $\mathbb{D}_P$ corresponding to Pearl's causal model P will have a special form: they will be *determinate* rules which have only literals in their heads. As has been shown in (**?**), the nonmonotonic semantics of a determinate causal theory contains only exact models that are generated by literals.

As has been mentioned by Pearl himself (in a footnote), the above representation ceases to be adequate for nonrecursive systems of structural equations that could have multiple solutions for the same value assignment of the exogenous variables. Such non-recursive systems has been studied in (**?**) and some other studies in structural equation models (especially those that deal with reversible causation). For the majority of such causal systems, the general nonmonotonic semantics of exact models, as well as the representation of (**?**), will still be appropriate. Moreover, such causal systems could still be viewed as *default* causal theories in which exogenous atoms play the role of defaults. However, the corresponding causal theories will include also plain assumptions (that are not defaults) that would be 'responsible' for creating additional exact models that are not generated by defaults only, and thereby would violate the condition of default grounding for default models from Definition 11 above.

An even more general approach to a causal representation of physical models described by usual mathematical equations was sketched in (**?**, Chapter 4), which was based on the use of non-determinate *axioms* of the form

$$\mathbf{t} \Rightarrow A,$$

where A provides a formal description of a mathematical equation in an appropriate (presumably first-order) logical language. The approach was illustrated on the use of the ideal gas law in causal discourse. It should be clear that such a law does not describe *directly* causal relations among the variables involved; the ideal gas law is only a "non-directional" mathematical equation

$$PV = kT$$

that determines functional relationships among the variables of pressure (P), volume (V), and temperature (T) of a fixed amount of gas. Still, such functional dependencies are not completely noncausal since they can be used to determine such causal relations in every actual instantiation of the model that could include also causal actions of heating (the gas chamber) and pushing (the piston of the chamber). Moreover, the whole procedure of using such axioms in causal reasoning turns out to be much similar to the methods of generating a *causal order* in a system described by a set of mathematical equations (**?**; **?**).

These are only some of the prospects for representing significant parts of physical discourse in causal reasoning (and thereby in knowledge representation and reasoning in AI), and they definitely seem worth further investigation.

6 Some Conclusions and Prospects

Initial versions of the causal calculus as described, e.g., in (**?**) and (**?**; **?**), were primarily focused on models that are worlds. This world-based semantics was even called the *causal* semantics in (**?**). Presumably, this primary focus was guided by the intended initial applications of the causal calculus in reasoning about action and change in AI. Anyway, it has been discovered already in an early (unpublished) paper of McCain and Turner *On relating causal theories to other formalisms* (1997) that causal rules $A \Rightarrow B$ under this world-based semantics can be interpreted as default rules of the form $:A/B$, namely premise-free default rules with a single justification. Moreover, this 'reverse' correspondence between causal calculus and default logic has been taken up and developed in (**?**) to a general correspondence between the world-based causal calculus and a particular kind of biconsequence relations.

Still, already in (**?**), it was shown that a more general nonmonotonic semantics for the causal calculus (that has been used as the main semantics in this paper) provides a natural representation of abductive reasoning. The present study complements this line of inquiry by showing that default logic can also be seen as a particular instantiation of the causal calculus under this more general semantics.

The distinction between a general semantics of the causal calculus and its world-based restriction has also an important logical aspect. Namely, the world-based nonmonotonic semantics admits a further rule for causal reasoning, Disjunction in the Antecedent, or reasoning by cases:

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

In fact, it was shown in (**?**) that the above inference rule and the corresponding system of *basic* causal inference constitutes, in effect, the 'internal logic' of causal reasoning in Pearl's causal models. Still, it was shown in the same paper that the rule Or becomes problematic in reasoning about actual causality, even though such a reasoning is restricted to (causal) worlds. Accordingly, a theory of actual causation presented in (**?**) was based, as the present paper, on regular inference. Moreover, the whole approach to causal reasoning in (**?**) was essentially based on viewing causal rules themselves as default assumptions in the sense of this study.

On a practical side, the suggested causal representation of default logic determines precise conditions under which causal reasoning can be 'simulated' by using corresponding deductive tools of default logic, and vice versa. In particular, it appears to provide a route by which the algorithms for computing extensions of default theories could be transferred to causal reasoning for computing both default and general exact models of causal theories.

In a more general perspective, however, the established correspondence between default logic and causal calculus can be seen as one more piece in a bigger picture that displays (significant parts of) nonmonotonic reasoning itself not just as a modern technical or practical invention of AI, but rather as an important contribution to a general theory of reasoning, a kind of reasoning that has its roots in antiquity. In this sense, a theory of causal reasoning both concurs and complements the modern theory of argumentation in AI. They both share a common assumption, or rather presumption, that nonmonotonic reasoning systems and formalisms should be viewed as instantiations of some general theory of reasoning, a theory that goes far beyond plain deductive reasoning. Much effort should still be invested, however, in establishing firm foundations for such a general theory, but the expected benefits make such an investment worth the effort, especially in our troubled times when nonmonotonic reasoning meets successful reasoning-free competitors in AI. It seems natural to argue that a unified, comprehensive and firmly grounded general theory of nonmonotonic reasoning could secure the role of knowledge representation and reasoning as an essential part of AI for a foreseeable future.

Acknowledgments

I would like to thank the anonymous reviewers for useful comments and suggestions.

References

Bochman, A., and Lifschitz, V. 2015. Pearl's causality in a logical setting. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, 1446–1452. AAAI Press.

Bochman, A. 2003. A logic for causal reasoning. In *IJCAI-03: Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 141–146. Morgan Kaufmann.

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

Bochman, A. 2005. *Explanatory Nonmonotonic Reasoning*. World Scientific.

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

Bochman, A. 2008. Default logic generalized and simplified. *Annals of Mathematics and Artificial Intelligence* 53:21–49.

Bochman, A. 2017. Argumentation, Nonmonotonic Reasoning and Logic. *FLAP* 4(8).

Bochman, A. 2018. On laws and counterfactuals in causal reasoning. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018*, 494–503. AAAI Press.

Bochman, A. 2021. *A Logical Theory of Causality*. MIT Press.

Bondarenko, A.; Dung, P. M.; Kowalski, R. A.; and Toni, F. 1997. An abstract, argumentation-theoretic framework for default reasoning. *Artificial Intelligence* 93:63–101.

Denecker, M.; Brewka, G.; and Strass, H. 2015. A formal theory of justifications. In *13th International Conference on Logic Programming and Non-monotonic Reasoning (LPNMR)*, 250–264.

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

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 Artificial Intelligence Research* 12:317–337.

Janhunen, T. 1999. On the intertranslatability of nonmonotonic logics. *Annals of Mathematics and Artificial Intelligence* 27:791–828.

Konolige, K. 1994. Using default and causal reasoning in diagnosis. *Annals of Mathematics and Artificial Intelligence* 11:97–135.

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

Marek, W., and Truszczyński, M. 1989. Relating autoepistemic and default logics. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning, KR'89*, 276–288. Morgan Kaufmann.

Marek, W., and Truszczyński, M. 1993. *Nonmonotonic Logic, Context-Dependent Reasoning*. Springer.

Marek, W.; Nerode, A.; and Remmel, J. 1990. A theory of nonmonotonic rule systems. *Annals of Mathematics and Artificial Intelligence* 1:241–273.

McCain, N., and Turner, H. 1997. Causal theories of action and change. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-97)*, 460–465.

Nayak, P. P. 1994. Causal approximations. *Artificial Intelligence* 70(1):277–334.

Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition. (1st ed. 2000).

Poole, D. 1988. A logical framework for default reasoning. *Artificial Intelligence* 36:27–47.

Poole, D. 1994. Representing diagnosis knowledge. *Annals of Mathematics and Artificial Intelligence* 11:33–50.

Reiter, R. 1980. A logic for default reasoning. *Artificial Intelligence* 13:81–132.

Simon, H. A. 1953. Causal ordering and identifiability. In Hood, W. C., and Koopmans, T. C., eds., *Studies in Econometric Method*. Wiley. 49–74.