A Causal Theory of Abduction

Alexander Bochman
Computer Science Department,
Holon Academic Institute of Technology, Israel
e-mail: bochmana@hait.ac.il

Abstract

The paper provides a uniform representation of abductive reasoning in the logical framework of causal inference relations. The representation covers in a single framework not only traditional, 'classical' forms of abduction, but also abductive reasoning in diagnosis, theories of actions and change, and abductive logic programming.

1 Introduction

Abduction is a kind of reasoning from facts to their explanations that is widely used now in many areas of AI, including diagnosis, action theories, truth maintenance, knowledge update and logic programming. In this study we are going to show that abduction can be given a uniform representation in terms of production and causal inference relations from [Bochman, 2004a]. Such inference relations provide a natural generalization of classical logic that allows for nonmonotonic reasoning. Accordingly, the suggested representation will clarify the role of causal reasoning in abduction, as well as the relation between abduction and nonmonotonic reasoning.

Causal considerations play an essential role in abduction. They determine, in particular, the very choice of abducibles, as well as the right form of descriptions and constraints (even in classical first-order representations). As has been shown already in [Darwiche and Pearl, 1994], system descriptions that do not respect the natural causal order of things can produce inadequate predictions and explanations.

The intimate connection between causation and abduction has become especially vivid in the so-called abductive approach to diagnosis (see, e.g., [Cox and Pietrzykowski, 1987; Poole, 1994; Konolige, 1994]). As has been acknowledged in these studies, reasoning about causes and effects should constitute a logical basis for diagnostic reasoning. Unfortunately, the absence of an adequate logical formalization for causal reasoning has relegated the latter to the role of an informal heuristic background, with classical logic serving as the representation language. This naturally raises the question whether classical logic is adequate for all kinds of abductive reasoning. We will give below certain grounds for a negative answer to this question.

In this study we suggest to base abductive reasoning entirely on causal descriptions. As we will see, the resulting formalism will subsume the 'classical' abductive reasoning. The formalism will provide us, however, with additional representation capabilities that will encompass important alternative forms of abduction, such as abductive reasoning in theories of actions and change and abductive logic programming. As a result, we obtain a generalized theory of abduction that covers in a single framework practically all kinds of abductive reasoning used in AI.

Our basic language will be the classical propositional language with the usual connectives $\{\wedge, \vee, \neg, \rightarrow, \mathbf{t}, \mathbf{f}\}$. $\vDash$ will denote the classical entailment. A Tarski consequence relation $\vdash$ in a classical language is *supraclassical* if it subsumes classical inference: $\vDash \subseteq \vdash$. By a *conditional theory* we will mean a set Δ of rules $A \vdash B$, where A, B are classical propositions. $\vdash_\Delta$ will denote the least supraclassical consequence relation containing Δ, and Cn_Δ its associated provability operator. As can be verified, $a \vdash_\Delta A$ holds iff A is derivable from a by the rules from Δ and the classical entailment.

A consequence relation is *classical* if it is supraclassical and satisfies the Deduction rule: if $a, A \vdash B$, then $a \vdash A \rightarrow B$. Classical consequence relation can be seen as a classical entailment with some additional, nonlogical axioms; it. satisfies already all the rules of classical inference.

2 Abductive systems and abductive semantics

We will describe first a formalization of standard abductive reasoning. This formalization will serve as a basis for our subsequent constructions and representations.

An *abductive system* is a pair $\mathbb{A} = (\mathrm{Cn}, \mathcal{A})$, where Cn is a supraclassical consequence relation, while $\mathcal{A}$ a distinguished set of propositions called *abducibles*. A set of abducibles $a \subseteq \mathcal{A}$ is *an explanation* of a proposition A, if $A \in \mathrm{Cn}(a)$.

In applications, the consequence relation Cn is usually given indirectly by a generating conditional theory Δ, in which case the corresponding abductive system can be defined as $(\mathrm{Cn}_\Delta, \mathcal{A})$. Many abductive frameworks also impose syntactic restrictions on the set of abducibles $\mathcal{A}$[1]. Thus, $\mathcal{A}$ is often restricted to a set of special atoms (e.g., those built from abnormality predicates ab), or to the corresponding set of literals. The restriction of this kind is not essential, however. Indeed, for any abducible proposition A we can introduce a

[1] Poole's Theorist [Poole, 1988a] being a notable exception.

new abducible propositional atom p_A, and add the equivalence $A \leftrightarrow p_A$ to the underlying theory. The new abductive system will have much the same properties.

An abductive system $(\mathrm{Cn}, \mathcal{A})$ will be called *classical* if Cn is a classical consequence relation. A classical abductive system can be safely equated with a pair $(\Sigma, \mathcal{A})$, where Σ is a set of classical propositions (the domain theory). An example of such a system in diagnosis is [de Kleer *et al.*, 1992], a descendant of the consistency-based approach of [Reiter, 1987].

In abductive systems, acceptance of propositions depends on existence of explanations, and consequently such systems sanction not only forward inferences determined by the consequence relation, but also backward inferences from facts to their explanations, and combinations of both. All these kinds of inference can be captured formally by considering only theories of Cn that are generated by the abducibles. This suggests the following notion:

Definition 2.1. The *abductive semantics* $\mathcal{S}_{\mathbb{A}}$ of an abductive system $\mathbb{A}$ is the set of theories $\{\mathrm{Cn}(a) \mid a \subseteq \mathcal{A}\}$.

By restricting the set of theories to theories generated by abducibles, we obtain a semantic framework containing more information. Generally speaking, all the information that can be discerned from the abductive semantics of an abductive system can be seen as abductively implied by the latter.

The information embodied in the abductive semantics can be made explicit using the associated Scott consequence relation, defined as follows[2]: for any sets b, c of propositions,

$$b \vdash_{\mathbb{A}} c \ \equiv\ (\forall a \subseteq \mathcal{A})(b \subseteq \mathrm{Cn}(a) \ \rightarrow\ c \cap \mathrm{Cn}(a) \neq \emptyset)$$

This consequence relation describes not only forward explanatory relations, but also abductive inferences from propositions to their explanations. Speaking generally, it describes the *explanatory closure*, or *completion*, of an abductive system, and thereby captures abduction by deduction (cf. [Console *et al.*, 1991; Konolige, 1992]).

Example. The following abductive system describes a variant of the well-known Pearl's example. Assume that an abductive system $\mathbb{A}$ is determined by the set Δ of rules

$$\mathit{Rained} \vdash \mathit{Grasswet} \qquad \mathit{Sprinkler} \vdash \mathit{Grasswet}$$
$$\mathit{Rained} \vdash \mathit{Streetwet},$$

and the set abducibles Rained, $\neg\mathit{Rained}$, $\mathit{Sprinkler}$, $\neg\mathit{Sprinkler}$, $\neg\mathit{Grassswet}$.

Since Rained and $\neg\mathit{Rained}$ are abducibles, Rained an independent (exogenous) parameter (and similarly for $\mathit{Sprinkler}$). However, since only $\neg\mathit{Grassswet}$ is an abducible, non-wet grass does not require explanation, but wet grass does. Thus, any theory of $\mathcal{S}_{\mathbb{A}}$ that contains $\mathit{Grasswet}$ should contain either Rained, or $\mathit{Sprinkler}$, and consequently we have

$$\mathit{Grasswet} \vdash_{\mathbb{A}} \mathit{Rained}, \mathit{Sprinkler}.$$

Similarly, $\mathit{Streetwet}$ implies in this sense both its only explanation Rained and a collateral effect $\mathit{Grasswet}$.

[2] A Tarski consequence relation of this kind has been used for the same purposes in [Lobo and Uzcátegui, 1997].

3 Production and causal inference

Production inference relations from [Bochman, 2004a] are based on rules of the form $A \Rightarrow B$ having an informal interpretation "*A produces, or explains, B*". Formally, a ***(regular)*** ***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 $A \Rightarrow B \wedge C$;

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

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

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

From a logical point of view, the most significant 'omission' of the above set is the absence of the reflexivity postulate $A \Rightarrow A$. It is precisely this feature of production inference that creates a possibility of nonmonotonic reasoning.

Production rules are extended to rules with sets of propositions in premises by stipulating that, for a set u of propositions, $u \Rightarrow A$ hold if $\bigwedge a \Rightarrow A$ for some finite $a \subseteq u$. $\mathcal{C}(u)$ will denote the set of propositions explained by u:

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

The production operator $\mathcal{C}$ plays the same role as the usual derivability operator for consequence relations. In particular, it is a monotonic operator, that is, $u \subseteq v$ implies $\mathcal{C}(u) \subseteq \mathcal{C}(v)$. Moreover, $\mathcal{C}$ is a continuous operator.

A (monotonic) semantics of production inference relations is described below.

Definition 3.1.
- A *bimodel* is a pair of consistent deductively closed sets. A *production semantics* is a set of bimodels. A production semantics $\mathcal{B}$ is *inclusive*, if $v \subseteq u$, for any bimodel (u, v) from $\mathcal{B}$.
- A production rule $A \Rightarrow B$ is *valid* in a production semantics $\mathcal{B}$ if, for any bimodel (u, v) from $\mathcal{B}$, $A \in u$ only if $B \in v$.

Regular production relations are strongly complete for the inclusive production semantics (see [Bochman, 2004a]).

By a *causal theory* we will mean an arbitrary set of production rules. For any causal theory Δ, we will denote by $\Rightarrow_\Delta$ the least production relation that includes Δ. Clearly, $\Rightarrow_\Delta$ is the set of all production rules that can be derived from Δ using the postulates for production relations.

3.1 Causal and quasi-classical inference

The following two special kinds of production inference relations will play an important role in what follows.

A production relation will be called *causal*, if it satisfies

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

and *quasi-classical*, if it is causal and satisfies

(Weak Deduction) If $A \Rightarrow B$, then $\mathbf{t} \Rightarrow (A \rightarrow B)$.

Causal production relations allow for reasoning by cases, and hence they can be seen as systems of objective reasoning about the world. Moreover, the relevant production rules can already be interpreted as *causal rules*, since they provide a natural formal representation of ordinary causal assertions. A useful fact about such inference relations is that any production rule is reducible to a set of clausal rules $\bigwedge l_i \Rightarrow \bigvee l_j$, where l_i, l_j are classical literals. In addition, any rule $A \Rightarrow B$ is equivalent to a pair of rules $A \wedge \neg B \Rightarrow \mathbf{f}$ and $A \wedge B \Rightarrow B$.

The rules $A \wedge B \Rightarrow B$ are *explanatory rules*. Though logically trivial, they play an important explanatory role in causal reasoning by saying that, if A holds, B is self-explanatory (and hence does not require explanation). On the other hand, the rule $A \wedge \neg B \Rightarrow \mathbf{f}$ is a *constraint* that does not have an explanatory content, but imposes a factual restriction $A \rightarrow B$ on the set of interpretations.

Quasi-classical production relations will be shown to characterize classical abductive reasoning. Weak Deduction is equivalent to the following postulate:

(CA) If $\neg A \Rightarrow \mathbf{f}$, then $A \Rightarrow A$.

The postulate asserts that any constraint A is also a self-explainable proposition. This partial collapse of the distinction between factual and explanatory information is actually a first symptom of the limitations of classical reasoning in representing abduction.

3.2 Nonmonotonic semantics

Production inference relations determine also a natural nonmonotonic semantics, and provide thereby a logical basis for a particular form of nonmonotonic reasoning.

Definition 3.2. A *general nonmonotonic semantics* of a production inference relation is the set of all its *exact theories*, that is, sets u of propositions such that $u = \mathcal{C}(u)$.

The general nonmonotonic semantics of a causal theory Δ will be identified with the nonmonotonic semantics of $\Rightarrow_\Delta$.

An exact theory describes an informational state that is closed with respect to the production rules and such that every proposition in it is *explained* by other propositions accepted in this state. Accordingly, they embody an *explanatory closure assumption*, according to which any accepted proposition should also have explanation for its acceptance.

The general nonmonotonic semantics for causal theories is indeed nonmonotonic in the sense that adding new rules to the production relation may lead to a nonmonotonic change of the associated semantics, and thereby of derived information. This happens even though production rules themselves are monotonic, since they satisfy Strengthening (the Antecedent).

Exact theories are fixed points of the production operator $\mathcal{C}$. Since the latter operator is monotonic, exact theories (and hence the nonmonotonic semantics) always exist.

The causal semantics

For a causal interpretation of production rules, it is natural to restrict the nonmonotonic semantics to worlds.

Definition 3.3. A *causal nonmonotonic semantics* of a production inference relation or a causal theory is the set of all its exact worlds.

The causal nonmonotonic semantics of causal theories coincides with the semantics suggested in [McCain and Turner, 1997]. Moreover, it has been shown in [Bochman, 2003] that causal inference relations constitute a maximal logic adequate for this kind of nonmonotonic semantics.

McCain and Turner have established an important connection between the nonmonotonic semantics of a causal theory and completion of the latter. A finite causal theory Δ is *definite*, if it consists of rules of the form $A \Rightarrow l$, where l is a literal or $\mathbf{f}$. A *completion* of such a theory is the set of all classical formulas

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

for any propositional atom p, plus the set $\{\neg A \mid A \Rightarrow \mathbf{f} \in \Delta\}$. Then the classical models of the completion precisely correspond to exact worlds of Δ (see [Giunchiglia *et al.*, 2004]).

The completion formulas embody two kinds of information. As (forward) implications from right to left, they contain the material implications corresponding to the causal rules from Δ. In addition, left-to-right implications state that a literal belongs to the model only if one of its causes is also in the model. These implications reflect the impact of causal descriptions using classical logical means. Note, in particular, that explanatory rules $A \wedge l \Rightarrow l$ produce trivial forward implications, but contribute additional explanations for occurring literals. In this sense, they play the same role as *weak causes* from [Poole, 1994], namely rules that cannot be used for prediction, but only for explanation of observations.

4 Production inference and abduction

In this section we will show that production inference provides a formal representation for abductive reasoning in abductive systems. To this end, we will extend the relevant notion of explanation and say that an arbitrary set u of propositions *explains* a proposition A in an abductive system, if A is explainable by the abducibles that are implied by u. The following definition is based on viewing this notion of explanation as a kind of production inference.

Definition 4.1. A *production inference relation associated with an abductive system* $\mathbb{A}$ is a production relation $\Rightarrow_{\mathbb{A}}$ determined by all bimodels of the form $(u, \mathrm{Cn}(u \cap \mathcal{A}))$, where u is a consistent theory of Cn.

We will assume that the set of abducibles $\mathcal{A}$ of an abductive system is closed with respect to conjunctions, that is, if A and B are abducibles, then $A \wedge B$ is also an abducible. Then the above production inference relation admits a very simple syntactic characterization. Namely, $A \Rightarrow_{\mathbb{A}} B$ holds if and only if A implies some abducible that explains B.

Lemma 4.1. *If* $\Rightarrow_{\mathbb{A}}$ *is a production inference relation associated with an abductive system* $\mathbb{A}$*, then*

$$A \Rightarrow_{\mathbb{A}} B \;\; \textit{iff} \;\; (\exists C \in \mathcal{A})(C \in \mathrm{Cn}(A) \;\&\; B \in \mathrm{Cn}(C))$$

As a consequence, we obtain that abducibles of an abductive system correspond precisely to 'reflexive' propositions of the associated production relation.

Corollary 4.2. *If* $\Rightarrow_{\mathbb{A}}$ *is a production inference relation associated with an abductive system* $\mathbb{A}$*, then* $C \Rightarrow_{\mathbb{A}} C$ *iff* C *is* Cn*-equivalent to an abducible.*

Due to this correspondence, reflexive (self-explanatory) propositions of a production relation can be seen as abducibles, and hence we introduce

Definition 4.2. A proposition A will be called an *abducible* of a production inference relation $\Rightarrow$, if $A \Rightarrow A$.

Production inference relations corresponding to abductive systems form a special class described in the next definition.

Definition 4.3. A regular production relation will be called *abductive* if it satisfies

(Abduction) If $B \Rightarrow C$, then $B \Rightarrow A \Rightarrow C$, for some abducible A.

Production inference in abductive production relations is always mediated by abducibles. The following lemma describes the corresponding nonmonotonic semantics.

Lemma 4.3. *Exact theories of an abductive production relation are precisely sets of propositions of the form $\mathcal{C}(u)$, where u is a set of abducibles.*

The next result establishes a correspondence between abductive production relations and abductive systems.

Theorem 4.4. *A production inference relation is abductive if and only if it is generated by an abductive system.*

Finally, we will show that the abductive semantics of an abductive system coincides with the nonmonotonic semantics of the associated abductive production relation.

Theorem 4.5. *If $\Rightarrow_{\mathbb{A}}$ is a production inference relation corresponding to an abductive system $\mathbb{A}$, then the abductive semantics of $\mathbb{A}$ coincides with the general nonmonotonic semantics of $\Rightarrow_{\mathbb{A}}$.*

Thus, abductive production relations under the general nonmonotonic semantics provide a faithful representation of abductive reasoning. Moreover, the representation gives a logical definition to abducibles as propositions having a certain logical property (namely reflexivity).

As has been shown in [Bochman, 2004a], any production relation includes a greatest abductive subrelation; moreover, in many regular situations (e.g., when the production relation is well-founded) the latter determines the same nonmonotonic semantics. Now, since abductive production relations correspond exactly to abductive systems, this means that the general nonmonotonic semantics of a production relation is usually describable by some abductive system, and vice versa.

4.1 Abduction in literal causal theories

Now we will show that a certain well-known class of abductive systems can be directly interpreted as causal theories. The description below will demonstrate, in effect, that the causal reading of abductive systems has long been present in the study of abduction and diagnosis.

By a *literal* inference rule we will mean a rule of the form $a \vdash l$, where l is a propositional literal, and a a set of literals. A conditional theory Δ will be called *literal* one, if it consists only of literal rules. Finally, an abductive system $\mathbb{A} = (\Delta, \mathcal{A})$ will be called *literal* one, if Δ is a literal conditional theory, and the set of abducibles $\mathcal{A}$ is also a set of literals.

The above simplified abductive framework has been extensively studied in the theory of diagnosis under the name 'causal theory' (see, e.g., [Console *et al.*, 1991; Konolige, 1992; 1994; Poole, 1994]). The name has a different meaning in our study, namely it denotes an arbitrary set of production rules. It will be shown, however, that the these two notions of a causal theory are closely related.

Recall that a set of rules can also be viewed as a causal theory in our sense. Moreover, it has been shown earlier that abducibles can be incorporated into causal theories by accepting corresponding explanatory rules $A \Rightarrow A$. Accordingly, for an abductive system $(\Delta, \mathcal{A})$, we will introduce a causal theory $\Delta_{\mathcal{A}}$ which is the union of Δ (viewed as a set of production rules) and the set $\{l \Rightarrow l \mid l \in \mathcal{A}\}$.

To begin with, the abductive semantics of $\mathbb{A}$ is included in the general nonmonotonic semantics of $\Delta_{\mathcal{A}}$.

Lemma 4.6. *Any theory $\mathrm{Cn}_{\Delta}(a)$, where $a \subseteq \mathcal{A}$, is an exact theory of $\Delta_{\mathcal{A}}$.*

However, the reverse inclusion does not hold, even in the literal case, and it is instructive to clarify the reasons why this happens. First, if the causal theory $\Delta_{\mathcal{A}}$ is not well-founded, it may have exact theories that are not generated by abducibles. Second, $\Delta_{\mathcal{A}}$ may create new abducibles of its own, if some of the propositions happen to be inter-derivable. Taking a simplest example, if we have that both $p \vdash q$ and $q \vdash p$ belong to Δ, then both p and q will be abducibles of $\Rightarrow_{\Delta_{\mathcal{A}}}$.

Both the above reasons for a discrepancy will disappear, however, if Δ is an *acyclic* theory. Actually, a restriction of this kind has been used extensively in the literature - see, e.g., [Pearl, 1988; Console *et al.*, 1991; Poole, 1994].

A *dependency graph* of a literal conditional theory Δ is the directed graph with literals as nodes, in which the arcs are pairs (l, m) of literals, for which Δ contains a rule $l, a \vdash m$. A conditional theory is *acyclic*, if its dependency graph does not contain infinite descending paths. In what follows, we will use, however, a weaker condition that will be sufficient for our purposes.

Definition 4.4. A literal abductive system $(\Delta, \mathcal{A})$ will be called *abductively well-founded*, if any infinite descending path in the dependency graph of Δ contains an abducible.

Clearly, any acyclic theory will also be abductively well-founded. The following result shows that in this case the causal theory $\Delta_{\mathcal{A}}$ captures the 'abductive content' of the source abductive system.

Theorem 4.7. *If $\mathbb{A}$ is an abductively well-founded literal abductive system, then the abductive semantics of $\mathbb{A}$ coincides with the nonmonotonic semantics of $\Delta_{\mathcal{A}}$.*

The above result shows that, from the perspective of abductive reasoning, literal conditional theories can be viewed directly as causal theories.

Example. (*continued*) The following causal theory corresponds to the (literal) conditional theory from the Pearl's example, discussed earlier.

$$Rained \Rightarrow Grasswet \quad Sprinkler \Rightarrow Grasswet$$
$$Rained \Rightarrow Streetwet$$
$$Rained \Rightarrow Rained \quad \neg Rained \Rightarrow \neg Rained$$
$$Sprinkler \Rightarrow Sprinkler \quad \neg Sprinkler \Rightarrow \neg Sprinkler$$
$$\neg Grasswet \Rightarrow \neg Grasswet \quad \neg Streetwet \Rightarrow \neg Streetwet$$

As follows from the above result, the general nonmonotonic semantics of this causal theory coincides with the abductive semantics of the source abductive system, and hence it determines the same abductive inferences.

5 Causal abduction

For 'objective' applications of abduction, such as diagnosis and logic programming, we have to consider the stronger *causal* nonmonotonic semantics of production inference. As we mentioned, causal inference relations constitute an adequate logic for this semantics. The corresponding kind of abductive systems is described in the next definition.

Definition 5.1. An abductive system $\mathbb{A} = (\mathrm{Cn}, \mathcal{A})$ will be called *$\mathcal{A}$-disjunctive* if $\mathcal{A}$ is closed with respect to disjunctions, and Cn satisfies the following two conditions, for any abducibles $A, A_1 \in \mathcal{A}$, and arbitrary B, C:

- If $B \vdash A$ and $C \vdash A$, then $B \vee C \vdash A$;
- If $A \vdash B$ and $A_1 \vdash B$, then $A \vee A_1 \vdash B$.[3]

$\mathcal{A}$-disjunctive systems are precisely abductive systems that generate causal production relations:

Theorem 5.1. *An abductive production relation is causal iff it is generated by an $\mathcal{A}$-disjunctive abductive system.*

In contrast, classical abductive reasoning corresponds in this sense to quasi-classical production inference.

Theorem 5.2. *An abductive production relation is quasi-classical iff it is generated by a classical abductive system.*

An important negative consequence from the above two results is that classical abductive systems are already inadequate for reasoning with respect to the causal nonmonotonic semantics. This conclusion is immediate from the fact that causal inference relations constitute a maximal logic for the latter, and hence any postulate added to causal inference will extend the set of admissible models beyond exact worlds.

In fact, the distinction between causal and classical abductive reasoning has appeared as a distinction between consistency-based and abductive approach to diagnosis. Traditionally, the difference between the two has been described as a difference between finding the set of faults consistent with observations versus finding faults that explain (that is, entail) observations. Further studies have shown, however, that a slight generalization of the consistency-based approach provides a representation also for explaining observations (see [de Kleer *et al.*, 1992]). On the other hand, it has been shown already in [Poole, 1988b] that that the consistency based diagnosis can be represented via a completion of an abductive theory. The real difference between the two approaches can be seen, however, as the difference between a fully classical description of diagnosis systems (as in [de Kleer *et al.*, 1992]) and their causal description (see, e.g., [Konolige, 1994; Poole, 1994]). The earlier abductive approach of [Console *et al.*, 1991] can also be viewed as implicitly causal, since it used a completion of the conditional base as way of solving the abductive task.

[3] Cf. the rule Ab-Or in [Lobo and Uzcátegui, 1997].

The framework of causal inference also provides syntactic means for differentiating between explaining observations and finding models consistent with observations. Namely, for an observation O, adding a constraint $\neg O \Rightarrow \mathbf{f}$ to a causal theory amounts to reducing the causal nonmonotonic semantics to exact worlds that explain O. But if we want only to check consistency of O with other data, we can add a rule $\mathbf{t} \Rightarrow O$. By the decomposition of causal rules, the latter is equivalent to the combination of $\neg O \Rightarrow \mathbf{f}$ and the explanatory rule $O \Rightarrow O$ that makes O an abducible. Accordingly, the observation O is exempted from the burden of explanation, and hence is checked only for consistency. Note, however, that precisely this distinction disappears in quasi-classical inference relations (see the postulate (CA) in Section 3.1).

5.1 Abduction in logic programming

Finally we will show that abduction in logic programming is also representable as a special case of the causal framework.

The role of abduction in logic programming is twofold (see [Kakas *et al.*, 1998] for an overview). First of all, logic programs themselves are representable as abductive systems in which negated atoms play the role of abducibles. In this sense, logic programs are inherently abductive, and abduction provides a representation for negation as failure.

Abductive logic programs are defined as pairs $(\Pi, \mathcal{A})$, where Π is a logic program, and $\mathcal{A}$ a set abducible atoms. A formalization of abductive reasoning in this setting is provided by the *generalized stable semantics* [Kakas and Mancarella, 1990]. According to the latter, an abductive explanation of a query q is a subset S of abducibles such that there exists a stable model of the program $\Pi \cup S$ that satisfies q.

It has been shown in [Inoue and Sakama, 1998], however, that abductive logic programs under the generalized stable semantics are reducible to general disjunctive logic programs under the stable semantics. The relevant transformation of abductive programs can be obtained simply by adding to Π the program rules $p, \mathbf{not}\, p \leftarrow$, for any abducible atom p from $\mathcal{A}$. This reduction has shown, in effect, that abductive programs have the same representation capabilities as general logic programs.

Now, general logic programs has been shown in [Bochman, 2004b] to be representable as causal theories. The translation for the stable semantics is obtained as follows. First, any program rule $c, \mathbf{not}\, d \leftarrow a, \mathbf{not}\, b$ is translated into a causal rule $d, \neg b \Rightarrow \wedge a \rightarrow \vee c$. Second, the resulting causal theory is augmented with the causal version of the Closed World Assumption stating that all negated atoms are abducibles:

Default Negation $\neg p \Rightarrow \neg p$, for any propositional atom p.

The causal nonmonotonic semantics of the resulting causal theory will correspond precisely to the stable semantics of the source logic program. Moreover, unlike known embedding of logic programs into other nonmonotonic formalisms, namely default and autoepistemic logics, the causal interpretation of logic programs turns out to be bi-directional in the sense that any causal theory is reducible to a general logic program.

Combining the above representation results, we immediately obtain a causal interpretation of abductive logic programs. Fortunately, under the causal translation of program rules, Inoue and Sakama's rules $p, \mathbf{not}\, p \leftarrow$ correspond to causal rules $p \Rightarrow p$ that make each such p an abducible of the resulting causal theory. Accordingly, for an abductive program $(\Pi, \mathcal{A})$, we define the causal theory $\Delta_{\Pi,\mathcal{A}}$ as the union $tr(\Pi) \cup \mathcal{A}^- \cup \mathcal{A}^+$, where $tr(\Pi)$ is the set of causal rules corresponding to the rules of Π, $\mathcal{A}^-$ is the set of rules $\neg p \Rightarrow \neg p$, for all atoms p, and $\mathcal{A}^+$ is the set of rules $p \Rightarrow p$, for all $p \in \mathcal{A}$. Then we obtain

Theorem 5.3. *The generalized stable semantics of an abductive program $(\Pi, \mathcal{A})$ coincides with the causal nonmonotonic semantics of $\Delta_{\Pi,\mathcal{A}}$.*

Thus, abductive logic programs also correspond to causal theories under the causal nonmonotonic semantics, subject to the Closed World Assumption.

6 Conclusions

It has been shown that the framework of production and causal inference provides a uniform logical basis for abductive reasoning. The suggested causal representation of abduction is syntax-independent in the sense that abducibles are defined not as syntactically designated propositions, but as propositions satisfying certain logical property in a causal system, namely reflexivity (self-explanation) $A \Rightarrow A$.

The results of this study indicate also that causal reasoning constitutes an essential ingredient, and even a pre-condition, of abduction. A truly general formalization of abduction in its current applications in AI can be achieved only by taking into account the causal picture of a situation or a system.

It seems reasonable to suppose that the suggested causal theory of abduction could be useful also in other applications of abduction in AI. Taking only one example, the causal interpretation of abduction in logic programming naturally provides a logical interpretation for a 'mixed' framework of Poole's Independent Choice Logic [Poole, 2000]. Without going into details, the latter system is representable uniformly as a causal theory in which atomic choices are abducibles.

References

[Bochman, 2003] A. Bochman. A logic for causal reasoning. In *Proceedings IJCAI'03*, Acapulco, 2003. Morgan Kaufmann.

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

[Bochman, 2004b] A. Bochman. A causal logic of logic programming. In D. Dubois, C. Welty, and M.-A. Williams, editors, *Proc. Ninth Conference on Principles of Knowledge Representation and Reasoning, KR'04*, pages 427–437, Whistler, 2004.

[Console *et al.*, 1991] L. Console, D. Theseider Dupre, and P. Torasso. On the relationship between abduction and deduction. *Journal of Logic and Computation*, 1:661–690, 1991.

[Cox and Pietrzykowski, 1987] P. T. Cox and T. Pietrzykowski. General diagnosis by abductive inference. In *Proc. IEEE Symposium on Logic Programming*, pages 183–189, 1987.

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

[de Kleer *et al.*, 1992] J. de Kleer, A. K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 52:197–222, 1992.

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

[Inoue and Sakama, 1998] K. Inoue and C. Sakama. Negation as failure in the head. *Journal of Logic Programming*, 35:39–78, 1998.

[Kakas and Mancarella, 1990] A. C. Kakas and P. Mancarella. Generalized stable models: A semantics for abduction. In *Proc. European Conf. on Artificial Intelligence, ECAI-90*, pages 385–391, Stockholm, 1990.

[Kakas *et al.*, 1998] A. C. Kakas, R. A. Kowalski, and F. Toni. The role of abduction in logic programming. In D. M. Gabbay, C. J. Hogger, and R. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 5, pages 235–324. Oxford UP, 1998.

[Konolige, 1992] K. Konolige. Abduction versus closure in causal theories. *Artificial Intelligence*, 53:255–272, 1992.

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

[Lobo and Uzcátegui, 1997] J. Lobo and C. Uzcátegui. Abductive consequence relations. *Artificial Intelligence*, 89:149–171, 1997.

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

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

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

[Poole, 1988b] D. Poole. Representing knowledge for logic-based diagnosis. In *Proc. Int. Conf. on Fifth Generation Computer Systems*, pages 1282–1290, Tokyo, 1988.

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

[Poole, 2000] D. Poole. Abducing through negation as failure: Stable models within the independent choice logic. *Journal of Logic Programming*, 44:5–35, 2000.

[Reiter, 1987] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.