

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 commonsense reasoning from facts to their explanations that is widely used now in many areas of AI, including diagnosis, action theories, truth maintenance, knowledge assimilation, database updates and logic programming. In this study we are going to show that this kind of reasoning can be given a uniform, thoroughly formal and syntax-independent representation in terms of production and causal inference relations introduced in [Boc03, Boc04a]. The basic idea of such inference relations amounts to dropping the Reflexivity postulate of classical inference. Such logics can also be viewed as a most natural and immediate generalization of classical logic that allow for nonmonotonic reasoning. Accordingly, the suggested representation of abductive reasoning in this framework should hopefully 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 [DP94], 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., [CP87, Dar95, Poo94, Kon92, Kon94]). 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 first-order logic serving as the main representation language. This naturally rises the question whether classical logic can always be used as an adequate and sufficiently general underlying logic for abductive reasoning and diagnosis. We will give in what follows certain grounds for a negative answer to this question.

The above considerations could serve as an initial justification for our suggestion to base abductive reasoning entirely on causal descriptions. As we will see, the resulting formalism will subsume the ‘classical’ abductive reasoning as a special case, and we will give precise conditions when it is appropriate. The formalism will provide us, however, with additional representation capabilities that will encompass important alternative forms of abduction, namely abductive reasoning suitable for theories of actions and change, and for abductive logic programming. As a result, we obtain a generalized theory of abduction that covers in a single formal framework practically all kinds of abductive reasoning studied in the AI literature.

* * *

We will assume that our basic language is a classical propositional language with the usual classical connectives and constants $\{\wedge, \vee, \neg, \rightarrow, \mathbf{t}, \mathbf{f}\}$. \models will stand for the classical entailment, with Th as the associated provability operator.

In what follows, we will use a few facts about consequence relations in a classical language (see, e.g., [Boc01] for a more detailed description). A Tarski consequence relation \vdash in a classical language is *supraclassical* if it subsumes classical inference, that is, $\models \subseteq \vdash$. Supraclassicality allows for replacement of classically equivalent formulas in premises and conclusions of the rules. In addition, it allows to replace sets of premises by their classical conjunctions: $a \vdash A$ becomes equivalent to $\bigwedge a \vdash A$. Accordingly, such consequence relations can be viewed as certain binary relations on the set of classical propositions.

By a *conditional theory* we will mean an arbitrary set Δ of rules of the form $A \vdash B$, where A, B are classical propositions. Then \vdash_{Δ} 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 if and only if A is derivable from a by the rules from Δ and the classical entailment.

A consequence relation is called *classical* if it is supraclassical and satisfies

Deduction If $a, A \vdash B$, then $a \vdash A \rightarrow B$.

Classical consequence relations satisfy already all the familiar rules of classical inference, such as Contraposition and Disjunction in the Antecedent. Moreover, they are intimately related to the classical entailment. Namely, any rule $A \vdash B$ of a classical consequence relation is equivalent to $\vdash A \rightarrow B$. Consequently, any classical consequence relation can be seen as a classical entailment with some set of propositions added as additional, nonlogical axioms. In this respect arbitrary supraclassical consequence relations allow additional freedom in that they permit the use of auxiliary nonlogical inference rules $A \vdash B$ that are not reducible to the corresponding material implications.

2 Abductive systems and abductive semantics

As our starting point, we will describe a formalization of standard abductive reasoning that will subsume the majority of abductive frameworks suggested in the literature. This formalization will serve as a basis for our subsequent constructions and representations.

A general *abductive system* can be defined as a pair $\mathbb{A} = (\text{Cn}, \mathcal{A})$, where Cn is a supraclassical consequence relation, while \mathcal{A} is a distinguished set of propositions called *abducibles*. A proposition A is *explainable* in an abductive system \mathbb{A} if there exists a consistent set of abducibles $a \subseteq \mathcal{A}$ such that $A \in \text{Cn}(a)$; in this case the set a is called *an explanation* of A . Clearly, A may have, in general, a number of different (and even incompatible) explanations.

In practical applications, the supraclassical consequence relation Cn is usually given indirectly by a generating conditional theory Δ , namely by a set of inference rules of the form $a \vdash A$. A corresponding abductive system can be represented in this case by a pair (Δ, \mathcal{A}) . However, given a conditional theory Δ , we can identify Cn with Cn_Δ - the least supraclassical consequence relation containing Δ . Accordingly, the pair (Δ, \mathcal{A}) can be faithfully represented by an abductive system $(\text{Cn}_\Delta, \mathcal{A})$ as defined above.

The majority of abductive frameworks also impose syntactic restrictions on the set of abducibles \mathcal{A} ¹. In most cases, \mathcal{A} is restricted to a set of special atoms (e.g., those built from abnormality predicates *ab*), or to the corresponding set of literals. It should be noted, however, that the restriction of this kind is not as drastic as it may seem. Indeed, for any abducible proposition A we can introduce a new abducible propositional atom p_A , and add the equivalence $A \leftrightarrow p_A$ to the underlying theory. The modified abductive framework will have much the same properties and solutions.

An abductive system (Cn, \mathcal{A}) will be called *classical* if Cn is a classical consequence relation. As we mentioned, a classical consequence relation can be seen as a classical entailment augmented with a set of nonlogical axioms. Consequently, a classical abductive system can be safely equated with a pair (Σ, \mathcal{A}) , where Σ is a set of classical propositions (the domain theory), and \mathcal{A} a set of abducibles. A good example of such a system in diagnosis is [dKMR92], a descendant of the consistency-based approach of [Rei87].

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

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

¹Poole's Theorist [Poo88a] being a notable exception.

Note that a consequence relation Cn is uniquely determined by the set of *all* its theories. Accordingly, by restricting the set of these 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.

It is known (see [Boc01]) that any set of theories determines a unique Scott (multiple-conclusion) consequence relation. Accordingly, in most cases, the information embodied in the abductive semantics can be made explicit by considering the associated Scott consequence relation.

Definition 2.2. An *abductive consequence relation* associated with an abductive system \mathbb{A} is a Scott consequence relation $\vdash_{\mathbb{A}}$ determined by the abductive semantics $\mathcal{S}_{\mathbb{A}}$.

The abductive consequence relation associated with an abductive system can be defined explicitly as follows: for any sets b, c of propositions,

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

By the above description, $b \vdash_{\mathbb{A}} c$ holds if any set a of abducibles that explains b explains also at least one proposition from c .² In other words, $b \vdash_{\mathbb{A}} c$ holds if any explanation of b is also an explanation of some $C \in c$.

The abductive consequence relation $\vdash_{\mathbb{A}}$ is an extension of Cn obtained by restricting the set of its theories. As a result, it describes not only forward explanatory relations, but also abductive inferences from propositions to their explanations. For example, if C and D are the only abducibles that imply A in an abductive system, then we will have $A \vdash_{\mathbb{A}} C, D$. Speaking more generally, the abductive consequence relation describes the *explanatory closure*, or *completion*, of an abductive system, and allows thereby to capture the abductive process by deductive means (see [CDT91, Kon92]). In particular, in such abductive consequence relations the task of abduction for a particular observation O amounts simply to filtering its theories with respect to O (see [Bar00]).

Example. The following abductive system describes a variant of the well-known example from [Pea87]. Assume that an abductive system \mathbb{A} is determined by the set Δ of rules

$$\begin{aligned} & \textit{Rained} \vdash \textit{Grasswet} \\ & \textit{Sprinkler} \vdash \textit{Grasswet} \\ & \textit{Rained} \vdash \textit{Streetwet}, \end{aligned}$$

and the set abducibles

$$\mathcal{A} = \{\textit{Rained}, \neg\textit{Rained}, \textit{Sprinkler}, \neg\textit{Sprinkler}, \neg\textit{Grasswet}\}.$$

By stipulating that both *Rained* and $\neg\textit{Rained}$ are abducibles, we make *Rained* an independent (exogenous) parameter (and similarly for *Sprinkler*).

²A Tarski consequence relation of this kind has been used for the same purposes in [LU97].

However, since only $\neg\textit{Grasswet}$ is an abducible, non-wet grass does not require explanation, but wet grass does. Thus, any theory of $\mathcal{S}_{\mathbb{A}}$ that contains *Grasswet* should contain either *Rained*, or *Sprinkler*, and consequently we have

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

Similarly, *Streetwet* implies in this sense both its only explanation *Rained* and a collateral effect *Grasswet*.

As a general conclusion, we can say that abductive reasoning with respect to an abductive system amounts to extending the corresponding consequence relation to the associated abductive consequence relation.

3 Production and causal inference

Production inference relations, introduced in [Boc04a], are based on rules of the form $A \Rightarrow B$ that hold among classical propositions. A general informal interpretation of such rules is “*A produces, or explains, B*”. Though driven by very different considerations and objectives, production inference relations have their origin in input-output logics from [MvdT00].

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 rules 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 much the same role as the usual derivability operator for consequence relations. In particular, it is a monotonic operator, that is, $u \subseteq v$ implies $\mathcal{C}(u) \subseteq \mathcal{C}(v)$. Actually, due to compactness, \mathcal{C} is even a continuous operator. Note also that $\mathcal{C}(u)$ is always a deductively closed set.

A set u of propositions is a *theory* of a production relation, if it is deductively closed, and $\mathcal{C}(u) \subseteq u$. Theories of a production relation are closed with respect

to its production rules, and they have much the same properties as ordinary theories of consequence relations.

A (monotonic) semantics of production inference relations can be given in terms of pairs of deductively closed theories called bimodels.

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 $B \in v$.

As has been shown in [Boc04a], regular production inference relations are strongly complete for the inclusive production semantics.

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

It is important to note that causal theories are just sets of inference rules on classical propositions. Accordingly, they can also be used for generating (supraclassical) consequence relations. This possibility will be exploited later.

3.1 Causal and quasi-classical inference

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

Definition 3.2. 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 already be seen as systems of *objective* production inference, namely as systems of reasoning about complete worlds. Moreover, production rules of such relations can already be interpreted as truly *causal rules*, since they provide a natural formal representation of ordinary causal assertions. Such inference relations have been introduced in [Boc03] and shown to provide a complete characterization for the reasoning with causal theories from [MT97].

A useful fact about such inference relations is that any production rule is reducible to a set of *clausal* rules of the form $\bigwedge l_i \Rightarrow \bigvee l_j$, where l_i, l_j are classical literals. Another important fact is that any causal rule $A \Rightarrow B$ is equivalent to a pair of rules $A \wedge \neg B \Rightarrow \mathbf{f}$ and $A \wedge B \Rightarrow B$.

Causal rules of the form $A \wedge B \Rightarrow B$ are logically trivial, but they play an important explanatory role in causal reasoning. Namely, they say that, in any causally explained interpretation in which A holds, we can freely accept B , since

it is self-explanatory in this context. Accordingly, such rules can be called *explanatory rules*. On the other hand, the rule $A \wedge \neg B \Rightarrow \mathbf{f}$ is a constraint that does not have an explanatory content, but plainly asserts a factual constraint on the set of possible interpretations, namely that the classical implication $A \rightarrow B$ should hold in them. This decomposition neatly delineates two kinds of information conveyed by causal rules. One is a factual information that constraints the set of admissible models, while the other is an explanatory information describing what propositions are caused (explainable) in such models.

It is interesting to observe that the above notion of an explanatory rule corresponds precisely to the notion of a *weak cause* suggested in [Poo94]. The correspondence can be discerned from the fact that, just as Poole's weak causes, explanatory rules cannot be used for prediction, but only for explanation of observations. A formal support for this claim can be obtained from the contribution of explanatory rules to the completion of a causal theory, described later.

Quasi-classical production relations will be shown below to characterize 'classical' abductive reasoning. Weak Deduction asserts that material implications corresponding to production rules can be seen as universally valid propositions in production inference. It can be shown that to be equivalent to the following rule:

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

The above rule asserts that any constraint A is also a self-explainable proposition. This partial collapse of the distinction between factual and explanatory information is actually responsible for the fact that we need syntactic means in order to preserve the distinction between abducibles and factual propositions in classical abductive systems.

The semantics for the above two kinds of production inference can be obtained by considering only bimodels of the form (α, β) , where α, β are worlds. Generalizing a bit, the corresponding production semantics can be defined as a *possible worlds model* $\mathbb{W} = (W, \mathcal{B}, V)$, where W is a set of possible worlds, \mathcal{B} a binary accessibility relation on W , and V a valuation function on worlds. Validity of productions can now be defined as follows:

Definition 3.3. A rule $A \Rightarrow B$ is *valid* in a possible worlds model (W, \mathcal{B}, V) if, for any $\alpha, \beta \in W$ such that $\alpha \mathcal{B} \beta$, if A holds in α , then B holds in β .

Theorem 3.1. [Boc04a] *A production inference relation is*

- *causal iff it has a quasi-reflexive possible worlds model.*
- *quasi-classical iff it has a reflexive possible worlds model.*

3.2 Nonmonotonic semantics

In the preceding sections, we have described a formalization and monotonic semantics for the logical systems of production inference. It turns out, however, that production inference relations determine also a natural nonmonotonic

semantics, and provide thereby a logical basis for a particular form of non-monotonic reasoning. Namely, the fact that the production operator \mathcal{C} is not reflexive creates an important distinction among theories of a production relation.

Definition 3.4. • A theory u of a production inference relation will be called *exact*, if $u = \mathcal{C}(u)$.

- A set u of propositions is an *exact theory of a causal theory* Δ , if it is an exact theory of \Rightarrow_{Δ} .

An exact theory describes an informational state in which every proposition is *explained* by other propositions accepted in this state. Accordingly, restricting our universe of discourse to exact theories amounts to imposing a kind of an *explanatory closure assumption* on a production relation. Namely, it amounts to requiring that any accepted proposition should also have reason, or explanation, for its acceptance. This suggests the following notion of a nonmonotonic semantics:

Definition 3.5. A *general nonmonotonic semantics* of a production inference relation or a causal theory is the set of all its exact theories.

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 non-monotonic change of the associated semantics, and thereby to a nonmonotonic change in the derived information. This happens even though production rules themselves are monotonic, since they satisfy Strengthening (the Antecedent).

Exact theories are precisely fixed points of the production operator \mathcal{C} . Since the latter operator is monotonic and continuous, exact theories (and hence the nonmonotonic semantics) always exist. Moreover, the general properties of monotonic operators immediately imply that any theory of a production relation contains a greatest exact theory. In addition, any exact theory is included in a maximal exact theory. Unfortunately, exact theories are not closed with respect to arbitrary intersections, and consequently a least exact theory containing a given set of propositions does not always exist.

3.2.1 The causal nonmonotonic semantics

If we concentrate on an objective understanding of production rules as causal rules valid for worlds, it is only natural to consider also the corresponding restriction of the general nonmonotonic semantics to exact theories that are worlds. In this case, the principle of explanatory closure can be justifiably called the *principle of universal causation* (see [Tur99]).

Definition 3.6. A *causal nonmonotonic semantics* of a production inference relation or a causal theory is the set of all its exact worlds.

Since the causal nonmonotonic semantics forms a subset of the general non-monotonic semantics, it produces, in general, a larger set of nonmonotonic consequences. Moreover, the causal semantics is just a set of worlds, so its logical

content is exhausted by the classical propositional theory that is uniquely associated with this set of worlds. Note, however, that, unlike the general nonmonotonic semantics, the causal nonmonotonic semantics is not guaranteed to exist.

The causal nonmonotonic semantics of causal theories coincides with the semantics suggested in [MT97]. Moreover, it has been shown in [Boc03] 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 will be called *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 of the form

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

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

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 Proposition 6 in [GLL⁺04]).

The completion formulas embody two kinds of information. As (forward) implications from right to left, they contain all the material implications corresponding to the causal rules from Δ . In addition, left-to-right implications state, in effect, that a literal belongs to the model only if one of its causes is also in the model. The latter implications reflect precisely the impact of the 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 [Poo94].

4 Production inference and abduction

In this section we will show that production inference relations, coupled with the general nonmonotonic semantics, provide a formal representation for abductive reasoning in abductive systems.

4.1 Abductive production inference

In order to translate abductive systems into the framework of production inference, we will slightly 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 . Then the set of propositions that are explainable by u will coincide with $\text{Cn}(\text{Cn}(u) \cap \mathcal{A})$.

Now, the main idea behind the following representation consists in viewing this latter notion of an explanation as a particular kind of production inference. In other words, u will be taken to produce A if it explains A in the above sense.

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 $(u, \text{Cn}(u \cap \mathcal{A}))$, 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 it turns out that 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*

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

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 $A \Rightarrow A$.

It turns out that production inference relations corresponding to abductive systems form a special class that is 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 .

As can be seen, production inference in abductive production relations is always mediated by abducibles. The following lemma provides a description of the nonmonotonic semantics of such production relations.

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 shows that abductive production relations are exactly production inference relations that are generated by abductive systems.

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

Finally, the following basic result shows that the abductive semantics of an abductive system coincides with the general 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 $\Rightarrow_{\mathbb{A}}$.*

Due to the above results, abductive production relations, coupled with the general nonmonotonic semantics, can be seen as a faithful logical representation of abductive reasoning. Notice that abductive production relations provide in this sense a syntax-independent description of abduction, since abducibles were *defined* as propositions having a certain logical property with respect to the production relation (namely reflexivity).

Example. (continued) The following causal theory determines the abductive production relation corresponding to the Pearl’s example, described earlier.

$$\begin{array}{l}
\text{Rained} \Rightarrow \text{Grasswet} \quad \text{Sprinkler} \Rightarrow \text{Grasswet} \quad \text{Rained} \Rightarrow \text{Streetwet} \\
\text{Rained} \Rightarrow \text{Rained} \quad \neg \text{Rained} \Rightarrow \neg \text{Rained} \\
\text{Sprinkler} \Rightarrow \text{Sprinkler} \quad \neg \text{Sprinkler} \Rightarrow \neg \text{Sprinkler} \\
\neg \text{Grasswet} \Rightarrow \neg \text{Grasswet} \quad \neg \text{Streetwet} \Rightarrow \neg \text{Streetwet}
\end{array}$$

As follows from the preceding results, 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.

As has been shown in [Boc04a], any production inference relation includes a unique greatest abductive subrelation; moreover, in many regular situations (for instance, when the production relation is well-founded) the latter subrelation determines the same nonmonotonic semantics. Now, since abductive production relations correspond exactly to abductive systems, this implies that in ordinary cases the general nonmonotonic semantics of a production relation is describable by some abductive system, and vice versa. As a general conclusion, however, we can say that production inference constitutes a proper generalization of abductive reasoning, a generalization that goes beyond well-foundedness.

4.2 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., [CDT91, Kon92, Kon94, Poo94]). The name has a different meaning in our study, namely

it denotes an arbitrary set of production rules. It will be shown, however, that 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 reflexive rules $A \Rightarrow A$. Accordingly, for an abductive system (Δ, \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, it is easy to verify that the abductive semantics of \mathbb{A} is included in the general nonmonotonic semantics of $\Delta_{\mathcal{A}}$.

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

However, the reverse inclusion in the above lemma does not hold, even in our present, literal case, and it is important to clarify the reasons why this happens. First of all, the causal theory $\Delta_{\mathcal{A}}$ is not well-founded, in general, so it may have exact theories that are not generated by abducibles. Second, even in the well-founded case, the causal theory $\Delta_{\mathcal{A}}$ 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* conditional theory. As a matter of fact, a restriction of this kind has been used extensively in the literature - see, e.g., [CDT91, Poo94]. Pearl's belief networks [Pea88] can also be seen as belonging to this class, because they are defined in terms of directed acyclic graphs of dependencies.

By a *dependency graph* of a literal conditional theory Δ we will mean the directed graph such that its nodes are the literals occurring in Δ , while the arcs are the pairs of literals (l, m) , for which Δ contains a rule of the form $l, a \vdash m$. As usual, a literal conditional theory Δ will be called *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 $\mathbb{A} = (\Delta, \mathcal{A})$ will be called *abductively well-founded*, if any infinite descending path in the dependency graph of Δ contains an abducible from \mathcal{A} .

An abductive system is abductively well-founded, if it does not have infinite descending chains of dependencies that consist of non-abducibles only. This implies, in particular, that non-abducibles do not form loops of dependencies. Clearly, any acyclic theory will also be abductively well-founded. Note also that we do not require that abducibles should not appear in heads of the rules, a condition often imposed on such abductive systems.

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.

5 Causal abduction

The abductive reasoning described in the preceding sections is still too general for many applications. The reason is that the general nonmonotonic semantics of production inference is largely epistemic, since it is based on exact theories that are in general incomplete. Consequently, such a semantics is too weak for ‘objective’ applications such as diagnosis or logic programming. For the latter, we should consider abductive reasoning that is based on the *causal* nonmonotonic semantics.

As we mentioned earlier, causal inference relations constitute a maximal logic suitable for the causal nonmonotonic semantics. It turns out that the corresponding form of abductive reasoning is determined by abductive systems described in the next definition.

Definition 5.1. An abductive system $\mathbb{A} = (\text{Cn}, \mathcal{A})$ 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 $A \in \text{Cn}(B)$ and $A \in \text{Cn}(C)$, then $A \in \text{Cn}(B \vee C)$;
- If $B \in \text{Cn}(A)$ and $B \in \text{Cn}(A_1)$, then $B \in \text{Cn}(A \vee A_1)$.³

The following result shows that \mathcal{A} -disjunctive systems are precisely abductive systems that generate causal production relations.

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

In contrast, the next result shows that classical abductive reasoning corresponds in this sense to quasi-classical production inference.

Theorem 5.2. *An abductive production relation is quasi-classical if and only if 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 additional postulate added to causal inference will extend the set of admissible models beyond exact worlds.

The distinction between causal and classical abductive reasoning can be illustrated by comparing two approaches to diagnosis, namely consistency-based and abductive approach. 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

³This rule corresponds to the rule Ab-Or in [LU97].

[dKMR92]). On the other hand, it has been shown already in [Poo88b] that that the consistency based diagnosis can be represented via a completion of an abductive theory. The real difference between the two approaches could be seen, however, as the difference between a fully classical description of diagnosis systems (as in [dKMR92]) and their causal description (see, e.g., [Kon94, Poo94]). The earlier abductive approach of [CDT91] can also be viewed as implicitly causal, since it used a completion of the source conditional base as way of obtaining solutions to the abductive task.

As a final remark, note that the framework of causal inference also provides syntactic tools for differentiating between explaining observations and finding models (or abducibles) consistent with observations. Namely, if O is an observation, then adding a factual constraint $\neg O \Rightarrow \mathbf{f}$ to a causal theory amounts to reducing the causal nonmonotonic semantics to exact worlds that explain O . On the other hand, 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 the same constraint $\neg O \Rightarrow \mathbf{f}$ 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. It should be noted, 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 [KKT98] for an overview). First of all, logic programs themselves are representable as abductive inference 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. This fact makes it only natural to use logic programming in abductive reasoning.

Just as general abductive systems, abductive logic programs are defined as pairs (Π, \mathcal{A}) , where Π is a logic program, while \mathcal{A} a set of propositional atoms called abducibles. A most influential formalization of abductive reasoning in logic programming is provided by the *generalized stable semantics* suggested in [KM90]. 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 turned out, however, that, due to the inherently abductive nature of logic programs, abductive logic programs under the generalized stable semantics are reducible to general disjunctive logic programs under a stable semantics. Namely, it has been shown in [IS98] that 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 (see also [SI00]).

Now, it has been shown in [Boc04b] that general logic programs are representable as causal theories under the causal nonmonotonic semantics. Thus, the relevant transformation for the stable semantics is obtained as follows. First, any general 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 $\neg p \Rightarrow \neg p$, where p is a propositional atom.

Then 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 interpretation of program rules, Inoue and Sakama's rules $p, \mathbf{not} p \leftarrow$ correspond precisely to causal rules $p \Rightarrow p$ that make each such p an abducible of the resulting causal theory. Accordingly, abductive logic programs corresponds precisely to causal theories under the causal nonmonotonic semantics, plus the Closed World Assumption.

It is interesting to note that the distinction between consistency-based and abductive explanation of observations in a causal setting, sketched at the end of the preceding section, corresponds precisely to the distinction made in [PEB94] in the framework of logic programs. Namely, a consistency-based diagnosis of an observation O can be obtained by adding the rule $O \leftarrow$ to the program, while its abductive diagnosis is achievable by adding O as an integrity constraint, namely by adding $\leftarrow \mathbf{not} O$.

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 and fully adequate account 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 (see [Poo00]). Without going into details, the latter system seems to be representable uniformly as a causal theory in which atomic choices play the role of abducibles.

References

- [Bar00] C. Baral. Abductive reasoning through filtering. *Artificial Intelligence*, 120:1–28, 2000.
- [Boc01] A. Bochman. *A Logical Theory of Nonmonotonic Inference and Belief Change*. Springer, 2001.
- [Boc03] A. Bochman. A logic for causal reasoning. In *Proceedings IJCAI’03*, Acapulco, 2003. Morgan Kaufmann.
- [Boc04a] A. Bochman. A causal approach to nonmonotonic reasoning. *Artificial Intelligence*, 160:105–143, 2004.
- [Boc04b] 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.
- [CDT91] L. Console, D. Theseider Dupre, and P. Torasso. On the relationship between abduction and deduction. *Journal of Logic and Computation*, 1:661–690, 1991.
- [CP87] P. T. Cox and T. Pietrzykowski. General diagnosis by abductive inference. In *Proc. IEEE Symposium on Logic Programming*, pages 183–189, 1987.
- [Dar95] A. Darwiche. Model-based diagnosis using causal networks. In *Proceedings Int. Joint Conf. on Artificial Intelligence, IJCAI-95*, pages 211–217, Montreal, 1995. Morgan Kaufmann.
- [dKMR92] J. de Kleer, A. K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 52:197–222, 1992.
- [DP94] A. Darwiche and J. Pearl. Symbolic causal networks. In *Proceedings AAAI’94*, pages 238–244, 1994.
- [GLL⁺04] E. Giunchiglia, J. Lee, V. Lifschitz, N. McCain, and H. Turner. Nonmonotonic causal theories. *Artificial Intelligence*, 153:49–104, 2004.
- [IS98] K. Inoue and C. Sakama. Negation as failure in the head. *Journal of Logic Programming*, 35:39–78, 1998.

- [KKT98] 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.
- [KM90] 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.
- [Kon92] K. Konolige. Abduction versus closure in causal theories. *Artificial Intelligence*, 53:255–272, 1992.
- [Kon94] K. Konolige. Using default and causal reasoning in diagnosis. *Annals of Mathematics and Artificial Intelligence*, 11:97–135, 1994.
- [LU97] J. Lobo and C. Uzcátegui. Abductive consequence relations. *Artificial Intelligence*, 89:149–171, 1997.
- [MT97] N. McCain and H. Turner. Causal theories of action and change. In *Proceedings AAAI-97*, pages 460–465, 1997.
- [MvdT00] D. Makinson and L. van der Torre. Input/Output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.
- [Pea87] J. Pearl. Embracing causality in formal reasoning. In *Proceedings AAAI-87*, pages 369–373, Seattle, 1987.
- [Pea88] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- [PEB94] C. Priest, K. Eshghi, and B. Bertolino. Consistency-based and abductive diagnoses as generalised stable models. *Annals of Mathematics and Artificial Intelligence*, 11:51–74, 1994.
- [Poo88a] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47, 1988.
- [Poo88b] D. Poole. Representing knowledge for logic-based diagnosis. In *Proc. Int. Conf. on Fifth Generation Computer Systems*, pages 1282–1290, Tokyo, 1988.
- [Poo94] D. Poole. Representing diagnosis knowledge. *Annals of Mathematics and Artificial Intelligence*, 11:33–50, 1994.
- [Poo00] D. Poole. Abducing through negation as failure: Stable models within the independent choice logic. *Journal of Logic Programming*, 44:5–35, 2000.
- [Rei87] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

- [SI00] C. Sakama and K. Inoue. Abductive logic programming and disjunctive logic programming: Their relationship and transferability. *Journal of Logic Programming*, 44:71–96, 2000.
- [Tur99] H. Turner. A logic of universal causation. *Artificial Intelligence*, 113:87–123, 1999.