Abduction is not Deduction-in-Reverse

MARTA CIALDEA MAYER, *Dipartimento di Informatica, Terza Università di Roma, via C. Segre 2, 00146 Roma, Italy.*
E-mail: cialdea@inf.uniroma3.it

FIORA PIRRI, *Dipartimento di Informatica e Sistemistica, Università di Roma 'La Sapienza', via Salaria 113, 00198 Roma, Italy.*
E-mail: pirri@dis.uniroma1.it

Abstract

Abduction is a topic that attracts much interest in AI and automated reasoning research. Different approaches have been devised, that give a formalized account of explanatory reasoning, propose methods to compute explanations, frame abduction in the context of logic programming. However, the logical nature of abduction is still far from being clear and different specifications of the key underlying concepts have been given, that make it difficult to speak of abduction as a single well-defined form of reasoning.

This work is a preliminary discussion on the logical nature of abductive reasoning, emphasizing the fundamental difference between abductive and deductive inference. Some logical properties of the inference to the "best explanation" are put forward and analyzed when the underlying logic is any extension of classical propositional logic (first order logic, modal logic) or a non monotonic system.

Keywords: abduction, ????

1 Introduction

Abduction is a natural form of reasoning in any scientific field where new knowledge may be gained by observation of real world facts. It is also a common form of everyday reasoning and its formalization has recently attracted much attention in the AI community. In fact, abduction plays an important role in intelligent problem solving tasks such as diagnosis, natural language interpretation, plan recognition and many others aspects yet not deeply investigated, like program debugging.

The first efforts to formalize explanation based reasoning come from the area of diagnostic problem solving, where two fundamental approaches have been devised: the consistency based approach pioneered by de Kleer [6] and formalized by Reiter [30] and the abductive approach introduced by Reggia[29], Cox and Pietrzykowsky [5] and Poole [28]. Poole [26] and successively Console and Torasso [4] have shown under which restrictions the two approaches compute the same explanations, and Konolige [18] generalized their results. The formulation proposed by Konolige is based on the definition of a causal theory, where explanations are conjunctions of atomic sentences belonging to a predefined set. Consistency based and abductive diagnosis are compared and generalized in the context of causal theories. A discussion on the difference between diagnosis and the logic of explanation is addressed in [24].

Another domain in which abduction has been widely considered is logic programming. One of the first formulations related to logic programming is presented in [28] where the system *Theorist* is introduced. More recently, Poole [27] proposes probabilistic Horn abduction as an extension of pure Prolog. Abduction has been studied in logic programming also to cope with non monotonic negation and SLDNF-resolution. Eshghi and Kowalski [9] point out that since abduction and deduction play a similar role, every logic program can be transformed into an abductive framework. Furthermore they propose an abduction-based procedure to compute negation by exploiting the stable model semantics of [13]. In [8], Dung showed that their procedure was not correct by exhibiting a counterexample. However on the trace of Eshghi and Kowalski many other works have exploited SLDNF-resolution and stable models, such as [31] and [17].

Another face of the abductive universe is represented by the pure logical account of abduction. Fall in this domain the proof-theoretical methods based on resolution [5, 7, 16, 22, 21], or on semantic tableaux and sequent calculi [2, 3]. Recently, in [1] Boutilier and Becher investigated also a formulation of abduction in terms of belief revision. They capture the notion of explanation through a conditional connective so as to deal with abduction as a revision process of the state of belief of an agent. Even if it differs in the aim and presentation, the idea relies on the account of explanations presented in Gärdenfors [11], where the connection between observations and explanations is evaluated relatively to an epistemic state.

Although important efforts have been made to give a satisfactory formalization of abductive reasoning, the logic of explanations has not yet received the deserved attention in the AI community and the logical nature of abduction is far from being clear. In the AI approach, the question is often settled by calling upon the duality between abduction and deduction and the attention has been mainly focused on the problem of computing abductive explanations.

A very first attempt to understand the logical nature of abduction has been accomplished by Zadrozny [34]. Zadrozny studies the invariance of plausible sets of ‘abductive rules’ formulated with respect to some specific domain in which abduction is performed, like anaphora resolution or diagnosis. Zadrozny’s proposal is the first, in the AI community, that highlights the deep difference between abduction and deduction. However, the inference rules that he proposes do not embody the notion of preference, that is still characterized as an external requirement at the meta-level.

This paper discusses, in an introductory and informal way, some basic requirements for a logic of explanations and gives an insight in the relation between abductive and non monotonic inference, often considered two aspects of the same coin and almost the same form of reasoning. The emphasis is on the fundamental difference between abductive and deductive inference when preference is treated at the logical level. Some logical properties of the inference to the “best explanation” are put forward and analyzed when the underlying logic is any extension of classical propositional logic (first order logic, propositional modal logic) or a non monotonic system.

1.1 Preference

The initial informal view of abduction relates it to the notion of explanation: α is abduced from β means that α is a good reason for the truth of β. The statement that abductive reasoning is the inference of α from the premisses $\alpha \rightarrow \beta$ and β betrays its primitive meaning, by flattening explanatory implication into material implication. The two notions cannot be confused; they enjoy different properties. For example, assume that A observes that her coffee is sweet enough; a good reason for it is that some spoon of sugar has been added to the coffee; however, it is not true that another reason for the coffee to be sweet is that some spoon of sugar has been added to the coffee and some spoon of salt has been added to the coffee. That is, monotonicity generally fails for explanatory implication. However, abduction cannot be reduced to a form of non monotonic reasoning. This appears, for instance, when noticing that another property which explanatory implication does not enjoy in general is reflexivity, not only because explaining a fact by itself is a rhetoric assertion, but mainly because simpler explanations may exist and, since Galileo, out of two hypotheses, the simpler is to be preferred.

The notion of preference between hypotheses is indeed the key problem of abduction and what makes it different from logical implication: not every assumption that implies an observed phenomenon explains it. Among such alternative hypotheses a preference relation is assumed such that only the most preferred ones are actually to be taken as explanations.

Preference criteria have been stated in many different ways, from absolutely general to very empirical ones. Peirce [23] says that 'all our knowledge may be said to rest upon *observed facts*' and that 'observed facts do not, in themselves, contain any practical knowledge. Such knowledge must involve addition to the facts observed'. Addition consists in preferring *simpler* hypotheses, where simplicity is intended as a qualitative property that is not reducible to a logical one.

Harman, in [14] argues that the *inference to the best explanation* is a judgment based on considerations

> such as which hypothesis is simpler, which is more plausible, which explains more which is less *as hoc*, and so forth (pp. 89).

He does not provide criteria to make such a choice. Thagard, arguing against the lack of methodological issues in Harman, gives in [32] precise criteria to make this choice:

> Inference to the best explanation is inference to the theory that best satisfies the criteria of *consilience* and *simplicity*, as well as a third: *analogy* (pp. 89)[1].

The notion of consilience can be summarized as follows: a theory T_1 is more consilient than a theory T_2 if the set of phenomena that T_1 explains is greater than the set of phenomena explained by T_2. Under this view, no logical approach can be taken; otherwise an inconsistent theory would be maximally consilient. Differently from the two preceding authors, Hempel gives a logical account for the theory of explanations, introducing in [15], the notion of *systematic power* on which basis explanatory theories are compared. The notion of preference, in fact, is based on that of *explanatory power* of a theory, namely an explanatory theory is to be preferred over another one

[1] Italics are ours.

whenever, together with the background theory, the first explains more than the latter. However the way preference is formally defined does not raise the same problems as in Harman.

In the AI literature, preference has been given two main criteria of choice, both based on semantic and syntactic considerations:

1. context-dependent and
2. context-independent.

The first one is adopted by Boutilier and Becher [1], whose approach can be formulated as follows. Given the logics $CTO4$ or CO and a CTO4 or CO-revision model of the agent's knowledge K, if the *explanandum* β is accepted (i.e. β is verified in some world of the model), then an *explanans* α is preferred if it is true in all the most plausible worlds in which also β is true. The conditional model of Boutilier and Becher has the advantage of allowing revision of a theory with which an interesting explanans may be inconsistent. However, under their criterion of preference, β itself is a preferred *explanans*, as well as any propositional tautology.

This view, furthermore, does not account for the asymmetry between explanations: sentences true in the same worlds of the same model behave equivalently, so that given two *explanantes* α and γ for the same observation β, if α, β and γ are true out of the same worlds, then γ is also an *explanans* for α. Many arguments have been given against the symmetry of explanations. See, for example, [33].

By contrast, the context-independent notion of preference is a purely logical relation between assertion. It has been widely adopted in the AI approach to abduction. For example, in the logic programming context, a clause α that subsumes β is preferred to β. In [20] preference is defined as 'being simpler', where a (propositional) formula α is simpler than β whenever the set of literals occurring in α is a subset of those occurring in β. In the case of abduction in full first order logic [21, 2], α is preferred to β whenever $\beta \rightarrow \alpha$ is valid. See also [5, 30, 7, 25, 34].

All the above context-independent views of preference can be summed up by saying that α is an *explanans* for β with respect to a background knowledge Θ iff

1. $\Theta \cup \{\alpha\} \models \beta$ and
2. α is preferred, i.e. for all α' such that $\Theta \cup \{\alpha'\} \models \beta$, if $\models \alpha \rightarrow \alpha'$ then $\models \alpha' \rightarrow \alpha$.

In such a view, preference between hypotheses is 'absolute' and purely logical in nature. There is no place for a theory of preference that describes how preference is to be measured in a given context.

Against the context-independent view of preference it can be argued that the request for a theory of preference relies on a crucial issue. Each domain has its laws of preference. If these laws are considered together with the agent's background knowledge then the properties of logical implication risk not to be faithful to the standard meaning of preference – as we have seen above, discussing context-dependent preference. Not considering them makes preference a purely logical issue completely independent of the domain the agent is dealing with.

Both views can be settled within the same framework if an abductive structure is defined as consisting of a couple of theories, Θ and Π, where Θ is the background theory expressing the agent's current knowledge of the world, and Π is the theory embodying the preference criteria governing the domain relative to the task that the agent has to perform.

In such a view, an *explanans* α for an *explanandum* β in the abductive structure $\langle \Theta, \Pi \rangle$ is such that:

- α explains β with respect to the background knowledge Θ (i.e. $\Theta, \alpha \models \beta$) and
- α is a preferred explanation with respect to the criteria Π, that is, for every γ such that $\Theta, \alpha \models \beta$, if $\Pi \models \alpha \rightarrow \gamma$, then $\Pi \models \gamma \rightarrow \alpha$.

For example, suppose $\Theta = \{q \wedge r \rightarrow p, s \rightarrow q\}$ and let p denote the phenomenon to be explained; if $\Pi = \{r \rightarrow s\}$ then $s \wedge r$ is preferred to $q \wedge r$. Clearly the theory Π may be empty. In that case, preference criteria are satisfied by the entailment relation *sic et simpliciter* and the context-independent account of preference is obtained. On the other hand, if $\Pi = \Theta$ then we fall into the context-dependent view.

The need for a theory of preference is compelling when looking for an explanation that is more plausible or more typical than the others, on the basis of some assumptions that are not purely logical in nature. For instance: IF THERE IS A LOT OF TRAFFIC ON THE HIGHWAY I EXPECT THAT THERE IS A CAR ACCIDENT, and such a typical explanation is to be preferred upon the alternative that an ELEPHANT JUST ESCAPED FROM THE CIRCUS IS MAKING A PERFORMANCE ON THE HIGHWAY. Between the two hypothesis, however, there is no logical relation that can lead to prefer one or the other; the preference has to be represented in separate theory.

When a preference relation is given, what kind of inference is really abduction? If its properties are analyzed, it appears that abduction is a kind of reasoning that is completely different from deduction.

Intuitively, the classical consequence relation determines a sort of measure on formulae such that left hand side formulae are stronger than right hand side formulae. Abduction needs the exact inverse measure for formulae: a correct premiss is one that allows as an *explanans* (given a background knowledge) a weaker formula than the *explanandum*. While in classical deductive reasoning weakening the consequence is allowed (notice that 'right weakening' is admitted also in non monotonic forms of reasoning), in abductive reasoning weakening the *explanandum* is not correct in general. Briefly, we can say that while non monotonic reasoning differs from classical deductive inference because of the rejection of left weakening, abduction differs from both forms of deductive reasoning because it disregards also right weakening.

In this work we address the problem of characterizing abduction by focusing on the key role of minimality and preference criteria, disregarding other albeit important concepts involved in a complete and fair account of abductive reasoning, such as relevance of the explanation, consistency with the background theory, acceptability of the *explanans*, etc.

In the following section, a definition of the abduction relation is given, that is parametric w.r.t. the preference relation, and some of its structural and logical properties are studied. Section 3 considers the relation between abduction and non monotonic inference, as well as some aspects of abduction in a non monotonic system.

2 Properties of abductive inference

In the same way as an abductive inference depends on the logic underlying the background theory [3], it also depends upon the properties of the preference relation that is used.

We assume thereafter that a reflexive relation on formulae is defined: a 'preference criterion', that allows two potential explanations to be compared in order to determine whether one of them is better than the other. An explanation is therefore accepted as a good one whenever it is minimal w.r.t. such a relation.

The preference relation can be expressed by means of a preference theory Π. In fact, Π defines the partial order on formulae $\sqsubseteq$, such that $\alpha \sqsubseteq \gamma$ iff $\Pi \models \gamma \rightarrow \alpha$. In the following we shall not refer explicitly to Π, but use the $\sqsubseteq$ notation instead, denoting a reflexive relation on formulae. It admits indeed greater generality. There are in fact interesting cases where preference is not transitive (see Section 3.2). The symbol $\sqsubset$ denotes strict preference: $\alpha \sqsubset \gamma$ iff $\alpha \sqsubseteq \gamma$ and $\gamma \not\sqsubseteq \alpha$.

A second parameter we are considering is the logic $\mathcal{L}$ that induces the logical consequence relation. We assume that $\mathcal{L}$ is either an extension of propositional classical logic, such as first order logic or modal logic, or the non-monotonic system **P** (see Section 3.1). This is to ensure the validity of some logical and structural properties that will be used in the sequel. We shall use the symbol $\vdash_{\mathcal{L}}$ for derivability in $\mathcal{L}$. We assume that the language of $\mathcal{L}$ includes the propositional letter $\top$ that is true in every interpretation.

Given a logic $\mathcal{L}$ and a preference relation $\sqsubseteq$ on formulae, the abductive three-place relation 'In the theory Θ, α is a good reason for β' will be denoted by $\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta$ and it is defined as follows:

$$(\star) \quad \Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta \quad \text{iff} \quad \begin{array}{l} \text{(i) } \Theta \vdash_{\mathcal{L}} \alpha \rightarrow \beta, \text{ and} \\ \text{(ii) for every formula } \gamma \text{ such that } \Theta \vdash_{\mathcal{L}} \gamma \rightarrow \beta\text{:} \\ \quad \text{if } \gamma \sqsubseteq \alpha, \text{ then } \alpha \sqsubseteq \gamma \end{array}$$

Any instance of $(\star)$ satisfies the following property:

$$\text{If } \Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta, \text{ then } \Theta \vdash_{\mathcal{L}} \alpha \rightarrow \beta \tag{2.1}$$

In other words, best explanations are indeed explanations.

Note that $(\star)$ admits context-dependent preference too, when $\alpha \sqsubseteq \beta$ is defined as $\Theta \vdash_{\mathcal{L}} \beta \rightarrow \alpha$. However, contextual preference makes inference to the best explanation trivial, in the sense that the only minimal explanation of β is β itself (modulo equivalence in Θ): if α is an explanation for β, then $\Theta \vdash_{\mathcal{L}} \alpha \rightarrow \beta$ and for all γ, if $\Theta \vdash_{\mathcal{L}} \gamma \rightarrow \beta$ and $\Theta \vdash_{\mathcal{L}} \alpha \rightarrow \gamma$, then $\Theta \vdash_{\mathcal{L}} \gamma \rightarrow \alpha$; now, clearly $\Theta \vdash_{\mathcal{L}} \beta \rightarrow \beta$ and, since $\Theta \vdash_{\mathcal{L}} \alpha \rightarrow \beta$, it follows $\Theta \vdash_{\mathcal{L}} \beta \rightarrow \alpha$, so that $\Theta \vdash_{\mathcal{L}} \alpha \equiv \beta$.

On the grounds of definition $(\star)$, structural properties of the abductive relation can be studied, possibly relating each of them to corresponding properties of the preference relation. Some of the properties enjoyed by the abductive relation are listed below, in the form of inference rules. The proof of their validity immediately follows from the validity of corresponding rules in the logic $\mathcal{L}$, together with simple considerations about the preference relation.

1. If for any α, $\top \sqsubseteq \alpha$, then

$$\Theta \vdash_{\mathcal{L}} \beta \text{ iff } \Theta \vdash_{\mathcal{L}} \top \rightsquigarrow \beta$$

The above rule is a weak form of $\rightsquigarrow$-introduction.

2. And-right:

$$\frac{\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta \ ; \ \Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \gamma}{\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta \wedge \gamma}$$

3. Modus Ponens:

$$\frac{\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta \rightarrow \gamma \ ; \ \Theta \vdash_{\mathcal{L}} \alpha \rightarrow \beta}{\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \gamma}$$

4. E-Cut:

$$\frac{\Theta \vdash_{\mathcal{L}} \alpha \ ; \ \Theta, \alpha \vdash_{\mathcal{L}} \beta \rightsquigarrow \gamma}{\Theta \vdash_{\mathcal{L}} \beta \rightsquigarrow \gamma}$$

5. E-monotonicity:

$$\frac{\Theta \vdash_{\mathcal{L}} \alpha \qquad \Theta \vdash_{\mathcal{L}} \beta \rightsquigarrow \gamma}{\Theta, \alpha \vdash_{\mathcal{L}} \beta \rightsquigarrow \gamma}$$

We show how to prove the validity of E-monotonicity when $\mathcal{L}$ is any extension of classical logic where the following weak form of Cut and Monotonicity hold:
$\Gamma \vdash_{\mathcal{L}} \alpha$ and $\Gamma, \alpha \vdash_{\mathcal{L}} \beta$ then $\Gamma \vdash_{\mathcal{L}} \beta$ (C)
$\Gamma \vdash_{\mathcal{L}} \alpha$ and $\Gamma \vdash_{\mathcal{L}} \beta$ then $\Gamma, \alpha \vdash_{\mathcal{L}} \beta$ (M)

Assume $\Theta \vdash_{\mathcal{L}} \beta \rightsquigarrow \gamma$, then also $\Theta \vdash_{\mathcal{L}} \beta \rightarrow \gamma$. If moreover $\Theta \vdash_{\mathcal{L}} \alpha$, then, by (M), $\Theta, \alpha \vdash_{\mathcal{L}} \beta \rightarrow \gamma$. If β is not minimal with this property, then there exists $\varphi \sqsubset \beta$ such that $\Theta, \alpha \vdash_{\mathcal{L}} \varphi \rightsquigarrow \gamma$. This fact, together with $\Theta \vdash_{\mathcal{L}} \alpha$, leads to conclude, by (C), that $\Theta \vdash_{\mathcal{L}} \varphi \rightsquigarrow \gamma$, thus contradicting $\Theta \vdash_{\mathcal{L}} \beta \rightsquigarrow \gamma$.
Observe that the above rules (C) and (M) are, indeed, Cautious Monotonicity and non monotonic Cut – see further – therefore E-monotonicity is valid in non monotonic abduction. The same holds for the other inference rules.

6. Left Logical Equivalence:
If for every α and β such that $\vdash_{\mathcal{L}} \alpha \equiv \beta$, $\varphi \sqsubseteq \beta$ iff $\varphi \sqsubseteq \alpha$ and $\beta \sqsubseteq \varphi$ iff $\alpha \sqsubseteq \varphi$, then

$$\frac{\vdash_{\mathcal{L}} \alpha \equiv \beta \ ; \ \Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \gamma}{\Theta \vdash_{\mathcal{L}} \beta \rightsquigarrow \gamma}$$

This rule is correct only if the ordering is preserved by substitution of logically equivalent formulae; this is not the case if, for example, the preference relation is based on some notion of bounded resource reasoning.

7. If $\sqsubseteq$ is a well-founded ordering relation, then for any Θ and β, there exists α such that $\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta$.
In general, abductive problems are not bound to have a solution (see [16, 21]). The above property relates the existence of solutions to the behaviour of the ordering relation. Its validity can be shown by considering that if $\sqsubseteq$ is well-founded, then every non-empty set of formulae has minimal elements. Moreover, for any Θ and β there exists α such that $\Theta \vdash_{\mathcal{L}} \alpha \rightarrow \beta$, so that $\{\gamma \mid \gamma \sqsubseteq \alpha \text{ and } \Theta \vdash_{\mathcal{L}} \gamma \rightarrow \beta\}$ has a minimal element φ, and $\Theta \vdash_{\mathcal{L}} \varphi \rightsquigarrow \beta$.

8. Reflexivity

$$\frac{\vdash_{\mathcal{L}} \alpha \rightarrow \alpha}{\vdash_{\mathcal{L}} \alpha \rightsquigarrow \alpha}$$

Observe that from the failure of monotonicity, from $\vdash_{\mathcal{L}} \alpha \rightsquigarrow \alpha$ we cannot get $\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \alpha$. The meaning of the above rule is that reflexivity holds only if the context is empty, that is we have no way to get a weaker explanation for α than α itself. The above rule is another weak form of $\rightsquigarrow$ introduction.

As explanatory implication is not logical consequence, many properties that generally hold for logical consequence are in general unacceptable for the abductive relation. Here follows a set of examples, that is is by no means intended to be exhaustive.

The failure of fundamental properties, such as reflexivity and transitivity, is paradigmatic of the logical structure of abduction as a completely different structure of reasoning than deduction.

1. Reflexivity, in the form:

$$\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \alpha$$

This property holds only if there is no γ such that $\Theta \vdash_{\mathcal{L}} \gamma \rightarrow \alpha$ and γ is strictly preferred to α, i.e. $\gamma \sqsubset \alpha$. Observe that the previous condition is, for instance, verified whenever α is a unit formula and $\sqsubseteq$ is the $\vdash_{\mathcal{L}}$ relation.

As a counterexample, consider propositional classical logic, $\Theta = \{p \rightarrow q\}$ and $\alpha = p \wedge q$. If $\alpha \sqsubseteq \beta$ is defined as $\beta \vdash_{\mathcal{L}} \alpha$, then $p \sqsubset (p \wedge q)$, so that reflexivity does not hold.

2. Transitivity, in the form:

$$\frac{\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta \ ; \ \Theta \vdash_{\mathcal{L}} \beta \rightsquigarrow \gamma}{\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \gamma}$$

A counterexample is given by $\Theta = \{p \rightarrow q, r \rightarrow s, p \rightarrow t, q \wedge s \rightarrow t\}$, $\alpha = p \wedge r$, $\beta = q \wedge s$, $\gamma = t$. Transitivity does not hold whenever $p \sqsubseteq (p \wedge r)$.

A common sense counterexample is given by the following situation: THERE WAS A CAR ACCIDENT because of THE RUSH HOUR AND A DRIVER THAT DID NOT COMPLY WITH THE ROAD SIGNALS and THERE IS TRAFFIC because of A CAR ACCIDENT; however, THE RUSH HOUR AND A DRIVER THAT DOES NOT COMPLY WITH THE ROAD SIGNALS is a redundant explanation for the observation THERE IS TRAFFIC; THE RUSH HOUR, in fact, is a simpler and sufficient reason for traffic. So, explanation is not carried on by the transitivity of material implication.

3. Monotonicity, in the form:

$$\frac{\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta}{\Theta, \gamma \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta}$$

As the simplest counterexample, let $\Theta = \emptyset$, $\alpha = \beta = \gamma = p$; if $\top \sqsubseteq p$, monotonicity does not hold.

4. Cut, in the form:

$$\frac{\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta \ ; \ \Theta, \beta \vdash_{\mathcal{L}} \gamma \rightsquigarrow \delta}{\Theta, \alpha \vdash_{\mathcal{L}} \gamma \rightsquigarrow \delta}$$

This property holds only if $\alpha = \top$ and $\Theta \vdash_{\mathcal{L}} \beta$, and this case corresponds in fact to the acceptable form of cut stated before (E-cut).

As a counterexample of the general form, consider $\Theta = \{p \wedge q \rightarrow r, p \wedge s \rightarrow t, r \wedge s \wedge u \rightarrow t\}, \alpha = p \wedge q, \beta = r, \gamma = s \wedge u, \delta = t$. Then, $\Theta \vdash_{\mathcal{L}} p \wedge q \leadsto r$ and $\Theta, r \vdash_{\mathcal{L}} s \wedge u \leadsto t$. But whenever $s \sqsubseteq s \wedge u$, it is not true that $\Theta, p \wedge q \vdash_{\mathcal{L}} s \wedge u \leadsto t$.

5. And-left:

$$\frac{\Theta \vdash_{\mathcal{L}} \alpha; \quad \Theta \vdash_{\mathcal{L}} \beta \leadsto \gamma}{\Theta \vdash_{\mathcal{L}} \alpha \wedge \beta \leadsto \gamma}$$

This rule may look reasonable because, when an abductive inference is performed, the preferred explanation β is added to the theory Θ and, if $\Theta \vdash \alpha$, the theories resulting from addition of β and $\alpha \wedge \beta$ are equivalent. But this argument assumes a contextual notion of preference, where $\alpha \sqsubseteq \beta$ iff $\Theta \vdash_{\mathcal{L}} \beta \rightarrow \alpha$.

Without this assumption, the rule And-right is not valid in general. The following counterexample can be considered: $\Theta = \{p, q \wedge s \rightarrow r\}$, $\alpha = p$, $\gamma = r$ and $\alpha \sqsubseteq \beta$ is defined by $\vdash_{\mathcal{L}} \beta \rightarrow \alpha$. Now, $q \wedge s$ is a minimal explanation for r, but $p \wedge q \wedge s$ is not minimal, because $\vdash_{\mathcal{L}} p \wedge q \wedge s \rightarrow q \wedge s$, but the converse is not true.

In the same example, in the context of Θ, $p \wedge q \wedge s$ and $q \wedge s$ are equivalent.

6. Right weakening, in the form:

$$\frac{\Theta \vdash_{\mathcal{L}} \alpha \leadsto \beta \ ; \ \vdash_{\mathcal{L}} \beta \rightarrow \gamma}{\Theta \vdash_{\mathcal{L}} \alpha \leadsto \gamma}$$

Let $\Theta = \emptyset$, $\alpha = \beta = p \wedge q$, and $\gamma = p$. If $p \sqsubseteq (p \wedge q)$, right weakening does not hold.

7. $\forall$-introduction and $\exists$-elimination:

$\forall$-I:

$$\frac{\Theta \vdash_{\mathcal{L}} \beta[t] \leadsto \gamma}{\Theta \vdash_{\mathcal{L}} \forall x \beta[x] \leadsto \gamma}$$

$\exists$-E:

$$\frac{\Theta \vdash_{\mathcal{L}} \exists x \beta[x] \leadsto \gamma}{\Theta \vdash_{\mathcal{L}} \beta[t] \leadsto \gamma}$$

Straightforward counterexamples to both rules are when γ is $\beta[t]$ and $\sqsubseteq$ is logical consequence.

The failure of monotonicity is shared by abductive and non monotonic reasoning, so that it has sometimes led to think that abduction is a specialized form of non monotonic reasoning. See, in this respect, Section 3.1.

The failure of right weakening is the best example of the deep difference between abduction and any form of deductive reasoning, both monotonic and non monotonic: it is the formal counterpart of what already noticed, that abduction does not allow to weaken the *explanandum* leaving the *explanans* unchanged. On the other hand, the failure of And-left shows that it is not possible, in general, to strengthen a preferred *explanans*, not even by addition of already derivable formulae.

3 Non Monotonic Logic

3.1 Abduction and Non Monotonic Logic

Abductive and Non Monotonic reasoning share a common important property, i.e. non monotonicity. In fact, enriching the background theory can obviously produce new hypotheses that explain the observed event, so that a given explanation may cease to be a minimal one (better hypotheses can be obtained). However, in spite of this apparent common ground, abduction and non monotonic reasoning are different forms of inference and, in particular, abduction is by no means a special form of non monotonic inference.

Addressing the relation between abduction and non monotonic reasoning and their combination, we refer to that part of the literature on non monotonic reasoning that has investigated on the finitistic inference operation $\mid\!\sim$ (mainly Gabbay [10], Kraus, Magidor & Lehmann [19] and Gärdenfors and Makinson [12]).

The basic (and common to all approaches) idea of non monotonic reasoning is to infer, from a given knowledge Θ of the world, more than it is deductively implied from Θ. The studies about the non monotonic inference relation $\mid\!\sim$ have been directed to regulate this strongest inference relation, by adding to the classical inference rules some new rules for $\mid\!\sim$, analogue to the classical ones. A great deal with non monotonic inference systems has been that of establishing a cumulative inference relation. In other words a relation satisfying both *cautious monotonicity* and *non monotonic cut*. The import of cumulativity is that whatever is already derivable from given information Θ, does not alter the inference process both when it is added and when it is dropped from Θ. In a cumulative logic, abduction is cumulative, since any inference operation on Θ does not change the set of possible *explanantes* of a given *explanandum*.

We assume, here, a relation $\mid\!\sim$ for non monotonic inference which is cumulative and furthermore satisfies (Or), that is the system of postulates **P** given in Kraus, Magidor and Lehmann [19], namely:

1. If $\alpha \vdash \gamma$, then $\alpha \mid\!\sim \gamma$, (Supraclassicality)
2. If $\vdash \alpha \equiv \beta$ and $\alpha \mid\!\sim \gamma$, then $\beta \mid\!\sim \gamma$ (Left Logical Equivalence)
3. If $\vdash \alpha \rightarrow \beta$ and $\gamma \mid\!\sim \alpha$, then $\gamma \mid\!\sim \beta$ (Right Weakening)
4. If $\alpha \mid\!\sim \beta$ and $\alpha \mid\!\sim \gamma$, then $\alpha \mid\!\sim \gamma \wedge \beta$ (And-right)
5. If $\alpha \mid\!\sim \beta$ and $\alpha \mid\!\sim \gamma$, then $\alpha \wedge \beta \mid\!\sim \gamma$ (Cautious Monotonicity)
6. If $\alpha \mid\!\sim \gamma$ and $\beta \mid\!\sim \gamma$, then $\alpha \vee \beta \mid\!\sim \gamma$ (Or)
7. If $\alpha \wedge \beta \mid\!\sim \gamma$, then $\alpha \mid\!\sim \beta \rightarrow \gamma$ (Conditionalization)

Notice that Supraclassicality implies:

8. $\alpha \mid\!\sim \alpha$ (Reflexivity)

Furthermore, the above rules imply *Cut*:

9. If $\alpha \mid\!\sim \beta$ and $\alpha \wedge \beta \mid\!\sim \gamma$, then $\alpha \mid\!\sim \gamma$ (Cut)

If we observe that any instance of the general definition of abduction satisfies:

If $\Theta \vdash_{\mathcal{L}} \alpha \rightsquigarrow \beta$ then $\Theta \vdash_{\mathcal{L}} \alpha \rightarrow \beta$

and that Supraclassicality is the opposite, a sort of duality is recognized between the two forms of reasoning. In fact, this is reflected by the intuitive meanings of non monotonic consequence and abductive explanation:

(nmr) $\alpha \mid\!\sim \beta$: if α, normally β, i.e. α is a *good enough* reason to believe β.

(abd) $\vdash \alpha \leadsto \beta$: α is *one of the best* possible reasons to explain β.

So in the first case α may be accepted even if β is not one of its classical consequences, in the second case that α classically entails β is not a sufficient condition to accept it as one of the 'best' explanations.

This fact, together with the observations in Section 2, allow us to conclude that, in spite of the apparent link between the too forms of reasoning, the relation of non monotonic inference with abduction is very loose. Non monotonic logic, in fact, strengthens the inference power of classical logic, while preserving its deductive nature.

3.2 Abduction in Non Monotonic logic

In this section we give just some hints on the meaning of a non monotonic abduction. Abduction is by itself an unsound form of reasoning. The inference of an hypothesis α that sanctions β does not ensures that more information on the current state of affair still validates α as a good explanation. In such a case, a revision process should decide of the conflict between the old hypothesis α and the new information. This is, however a process that does not strictly attain the formalization of abduction, or at least the formalization of the general definition.

Boutilier has addressed the formulation of abduction and belief revision as a whole process (see [1]). He argues that even explanations should admit exceptions. In effect this is what normally happens in common sense explanation. A MATCH IS LIT is explained by IT HAS BEEN STRUCK, but if the match is wet, this is no longer true. In any explanation we implicitly mean that all the rest is fixed, or better, that any other event not causally connected with the phenomenon has been excluded. As van Fraassen says, in the explanation a tacit *ceteris paribus* clause *all else had been the same* is included.

So, in what consists abduction when the underlying logic is non monotonic? In order to address this point we need to relate the abductive relation between a theory Θ and a pair of formulae to the notion of proof in the non monotonic system given in [19]. Let **K** be a set of non monotonic axioms (of the form $\alpha \mid\!\sim \beta$) and Θ a finite set of propositional formulae; we shall identify a finite set of formulae with the conjunction of the formulae in the set. We say that $\Theta \mid\!\sim \alpha \leadsto \beta$ (in **K**) if $\Theta \mid\!\sim \alpha \rightarrow \beta$ is preferentially entailed by **K** – i.e. it can be derived from **K** in the system **P** – and α is minimal with such a property; in other words the set **K** is considered fixed.

Now, from a formal point of view, we observe that abduction based on non monotonic logic is still shaped in agreement with the observations in Section 2: the same inference rules remain valid and no new rule is validated. In fact, being the abductive rules a subset of the corresponding non monotonic ones, they are all preserved.

However, some observations are in order.

- Abduction in non monotonic reasoning does not satisfy monotonicity for a supplementary reason: assuming that $\Theta \mid\!\sim \alpha \leadsto \beta$, not only can $\Theta, \gamma \mid\!\sim \alpha \leadsto \beta$ be false because α ceases to be minimal, but also because $\Theta, \gamma \mid\!\sim \alpha \rightarrow \beta$ may be no longer true.

  The classical example of birds and penguins can be used to illustrate this point. For simplicity, we assume here that $\sqsubseteq$ is monotonic logical consequence. Let **K** consist of the following axioms:

$$p|\!\sim b \qquad p|\!\sim \neg f \qquad b|\!\sim f$$

and $\Theta = \emptyset$. Let f be the *explanandum*. From $\mathbf{K}$, by Conditionalization, we get $|\!\sim b \rightarrow f$. Since $|\!\sim f$ is not derivable, b is minimal and $|\!\sim b \leadsto f$. Observe that by Cautious monotonicity one also gets $p \wedge b|\!\sim \neg f$ but clearly it is not true that $|\!\sim p \wedge b \leadsto \neg f$; in fact $|\!\sim p \leadsto \neg f$.

Now, if the background theory is enriched with p, then b is no more an abductive explanation of f, i.e. $p|\!\sim b \leadsto f$ is false. And this is not because there is a better explanation than b, but because b no longer entails non monotonically f, i.e. $p|\!\sim b \rightarrow f$ fails to be true.

An example with non empty theory can be formulated as follows. Let $\Theta = \{p \wedge q \rightarrow r\}$ and $\mathbf{K} = \{p|\!\sim q\}$. Then $\Theta|\!\sim p \rightarrow r$ is derivable in $\mathbf{K}$, while $\Theta|\!\sim r$ is not, therefore $\Theta|\!\sim p \leadsto r$. However, $\Theta \cup \{\neg q\} \not|\!\sim p \leadsto r$. In fact $\Theta \cup \{\neg q\}|\!\sim p \rightarrow r$ is not preferentially entailed by $p|\!\sim q$.

- If the preference relation itself is defined in terms of non monotonic inference, i.e. if $\alpha \sqsubseteq \gamma$ iff $\gamma|\!\sim \alpha$ (in a given non monotonic preference theory Π, that may or may not be equal to $\mathbf{K}$,) then we obtain an example of a preference relation that is not transitive.
- Assume the set $\mathbf{K}$ of non-monotonic axioms is fixed. Abductive Supraclassicality (ANM-Supraclassicality):

$$\Theta \vdash \alpha \leadsto \beta \text{ implies } \Theta|\!\sim \alpha \leadsto \beta$$

does not hold. In fact, in order to establish that $\Theta|\!\sim \alpha \leadsto \beta$, the hypothesis α has to be compared with all the elements in $\{\gamma \mid \Theta|\!\sim \gamma \rightarrow \beta\}$, that is obviously a superset of $\{\gamma \mid \Theta \vdash \gamma \rightarrow \beta\}$

4 Concluding remarks

The discussion in this work shows that a faithful characterization of abduction cannot be hastily settled by saying that abduction is just a calculus exploiting the duality

$$\Theta, \alpha \vdash \beta \text{ iff } \Theta, \neg\beta \vdash \neg\alpha$$

and the existence of generative deduction methods. Abduction is a peculiar form of inference, that deserves attention as a stand-alone logic. The main issue, at which we are presently working, is the formulation of a suitable semantics, on whose grounds the meaning of the inference rules can be judged, as well as their completeness.

Acknowledgements

This work has been supported by the ESPRIT Basic Research Project MEDLAR II and the ASI project 'Stazione di lavoro per la generazione interattiva di piani per sistemi spaziali complessi'.

References

[1] C. Boutilier and V. Becher. Abduction as belief revision. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93)*, pages 642–648, 1993.

[2] M. Cialdea Mayer and F. Pirri. First order abduction via tableau and sequent calculi. *Bulletin of the IGPL* 1(1):99–117, 1993.

[3] M. Cialdea Mayer and F. Pirri. Propositional abduction in modal logic. *Journal of the IGPL* 3(6):907–919, 1995.

[4] L. Console and P. Torasso. A spectrum of logical definitions of model-based diagnosis. *Computational Intelligence* 7:133–141, 1991.

[5] P. T. Cox and T. Pietrzykowski. Causes for events: their computation and applications. In *Proceedings of the Eighth Conference on Automated Deduction (CADE-86)*, pages 608–621, 1986.

[6] J. de Kleer and C. B. Williams. Diagnosing multiple faults. *Artificial Intelligence* 32:97–130, 1987.

[7] R. Demolombe and L. Fariñas del Cerro. An inference rule for hypothesis generation. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 152–157, 1991.

[8] P.M. Dung. Negation as hypotheses: an abductive foundation for logic programming. In *International Conference on Logic Programming*, pages 3–18, 1991.

[9] K. Eshghi and R.A. Kowalski. Abduction compared with negation by failure. In *Proceedings of the Sixth International Conference on Logic Programming (ICLP-89)*, pages 234–254, 1989.

[10] D. M. Gabbay. Theoretical foundations for non monotonic reasoning in expert systems. In K. Apt, editor, *Logic and Models of concurrent systems*, pages 160–174. Springer-Verlag, 1985.

[11] P. Gärdenfors. *Knowledge in Flux*. Mit Press, Bradford Book, 1988.

[12] P. Gärdenfors and D. Makinson. Non monotonic inference based on expectations. *Artificial Intelligence Journal* 65:197–246, 1994.

[13] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In *Proceedings of the Fifth International Conference on Logic Programming (ICLP-88)*, pages 230–237, 1988.

[14] G.H. Harman. The inference to the best explanation. *Philosophical Review* 74:88–95, 1978.

[15] C.G. Hempel and Oppenheim. Studies in the logic of explanations. *Philosophy of Science* 15:135–175, 1948.

[16] K. Inoue. Linear resolution for consequence finding. *Artificial Intelligence Journal* 56:301–353, 1992.

[17] A. C. Kakas and Mancarella P. Generalized stable models: a semantics for abduction. In *Proceedings of the Ninth European Conference on Artificial Intelligence (ECAI-90)*, pages 385–391, 1990.

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

[19] S. Kraus, D. Lehmann, and M. Magidor. Non monotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence Journal* 44:167–208, 1990.

[20] H. J. Levesque. A knowledge-level account of abduction. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 1061–1067, 1989.

[21] P. Marquis. Extending abduction from propositional to first-order logic. In *Proceedings of the First International Workshop on Fundamentals of Artificial Intelligence Research (FAIR '91)*, pages 141–155, 1991. Springer-Verlag, LNAI 535.

[22] C. G. Morgan. Hypothesis generation by machine. *Artificial Intelligence Journal* 2:179–187, 1971.

[23] C.S. Peirce. Abduction and induction. In P. Owen, editor, *Classics in Logic*, pages 331–352. Academic Press, 1962.

[24] F. Pirri. *Abduction*. PhD thesis, Université Paris VI, 1995.

[25] D. Poole. Explanation and prediction: an architecture for default and abductive reasoning. *Computational Intelligence* 5:97–110, 1989.

[26] D. Poole. Normality and faults in logic-based diagnosis. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 1304–1310, 1989.

[27] D. Poole. Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence Journal* 64:81–131, 1993.

[28] D. Poole, R. Goebel, and R. Aleliunas. Theorist: a logical reasoning system for default and diagnosis. In N. Cercone and G. Mc Calla, editors, *The Knowledge Frontier*, pages 331–352. Springer Verlag, 1987.

[29] J.A. Reggia, D.S. Nau, and Y. Wang. Diagnostic expert systems based on a set covering model. *International Journal on Man Machine Studies* 19:437–460, 1983.

[30] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence Journal* 32:57–96, 1987.

[31] M. Shanahan. Prediction is deduction but explanation is abduction. In *Proceedings of the 8th National Conference on Artificial Intelligence*, pages 1055–1060, 1990.

[32] P. Thagard. The best explanation: criteria for theory choice. *Journal of Philosophy* 75:76–92, 1978.

[33] B. van Fraassen. *The Scientific Image*. Clarendon press, Oxford, 1980.

[34] W. Zadrozny. On rules of abduction. *Annals of Mathematics and Artificial Intelligence* 9:387–419, 1993.

Received 31 August 1994. Revised 24 October 1995