ELSEVIER

Artificial Intelligence 111 (1999) 131–169

Artificial Intelligence

www.elsevier.com/locate/artint

Jumping to explanations versus jumping to conclusions

Ramón Pino-Pérez [a,*], Carlos Uzcátegui [b,1]

[a] *Lab. d'Informatique Fondamentale de Lille, U.A. 369 du CNRS, Université de Lille I, 59655 Villeneuve d'Ascq, France*

[b] *Departamento de Matemáticas, Facultad de Ciencias, Universidad de Los Andes, Mérida 5101, Venezuela*

Received 24 November 1997; received in revised form 11 March 1999

Abstract

Abduction is usually defined as the process of inferring the best explanation of an observation. There are many information processing operations that can be viewed as a search for an explanation. For instance, diagnosis, natural language interpretation and plan recognition. This paper is concerned about the following aspects of abduction: (i) what are the logical properties of abduction when it is regarded as a form of inference? and (ii) how close is abduction to reversed deduction?

In the logic-based approach to abduction, the background theory is given by a consistent set of formulas Σ. The notion of an explanation is defined by saying that a formula γ (consistent with Σ) is an explanation of α if $\Sigma \cup \{\gamma\} \vdash \alpha$. An explanatory relation is a binary relation $\rhd$ among formulas where the intended meaning of $\alpha \rhd \gamma$ is "γ is a preferred explanation of α". To each explanatory relation is associated a consequence relation $\mid\!\sim_{ab}$ defined as follows: $\alpha \mid\!\sim_{ab} \beta$ if $\Sigma \cup \{\gamma\} \vdash \beta$ for each γ such that $\alpha \rhd \gamma$.

The study of the logical properties of explanatory reasoning is approached by a systematic analysis of $\mid\!\sim_{ab}$. We show that there are rationality postulates for abduction (i.e., constraints on the explanatory relation $\rhd$) that are, in a very precise sense, equivalent to rationality postulates (in the Krauss–Lehmann–Magidor tradition) for nonmonotonic reasoning (i.e., for the relation $\mid\!\sim_{ab}$). This tight correspondence between postulates for explanatory reasoning and nonmonotonic reasoning will make apparent a strong duality between these two forms of inference. Isolating the postulates and showing this duality are the main contributions of the paper. We introduce the notion of a causal explanatory relation and show its close connection with reversed nonmonotonic reasoning. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Abduction; Explanatory and nonmonotonic reasoning; Nonmonotonic consequence relations

* Corresponding author. Email: pino@lifl.fr.

[1] Email: uzca@ciens.ula.ve.

PII: S0004-3702(99)00038-7

1. Introduction

Abduction is usually defined as the process of inferring the best explanation of an observation. There are many information processing operations that can be viewed as a search for an explanation, and thus, as operations that perform some form of abduction.

(a) Diagnosis is the typical example of abduction. When a system (an electrical circuit, a trade market or something as complex as a living being) is ill-functioning or not functioning as expected, we seek for explanations that will help to return the system to its normal state. If there is more than one explanation, usually some relevance or simplicity criterion is invoked to guide the selection of the best explanation.

(b) We might need to explain an observation (input) α in order to give a meaning to it, because α itself is just a string of symbols. For instance, when reading a text we come across a word α that we do not know, we look up in a dictionary to give a meaning to it. If α has several senses, we select one of them according to the context.

(c) We can also use abduction when trying to make a plan to achieve a goal or to decide how to continue an activity. For example, in order to decide what to do after an experiment is made (maybe to confirm or disprove a conjecture), the output data has to be analyzed and then, in the best case, it will be explained by the background theory.

A traditional model of abductive reasoning assumes a deductive relationship between the explanandum (or fact to be explained) and its explanations. The basic idea is to model abduction as reversed deduction plus some additional conditions. In this logic based approach to abduction, the background theory is given by a consistent set of formulas (which will be denoted by Σ) and a formula γ is said to be an explanation of α (with respect to Σ) if $\Sigma \cup \{\gamma\}$ entails α. To avoid trivial explanations it is also required that an explanation has to be a formula consistent with Σ. Since abduction is the process of inferring the "best" explanation, this notion of explanation captures only possible or candidate explanations of α. Thus some additional conditions are needed to define the key notion of "preferred explanations". We are concerned about the following three aspects of abduction:

(i) what are the logical properties of abduction when it is regarded as a form of inference?,

(ii) how close is abduction to reversed deduction?, and

(iii) since preference criteria for selecting explanations are so fundamental to abduction, how is (i) and (ii) related to the selection mechanism?

Let us see these three aspects separately.

(i) Several people have studied the logical properties of abductive reasoning: Zadrozny [20], Flach [5], Cialdea and Pirri [4] and Aliseda [2]. They have approached the problem by isolating rationality postulates or rules that abductive reasoning should conform to. As Zadronzny put it, abduction is an inference process that preserves sets of explanations. The structural properties we are looking for should provide a clear picture of the peculiar features that truly makes abduction a form of logical inference. The following are two basic questions related to this aspect:

(a) How much a change of an observation affects its explanations? For instance, suppose that γ is a preferred explanation of $\alpha \wedge \beta$. Should γ be considered also a preferred explanation of α? Another example, if γ is a preferred explanation of α and also of β, is γ a preferred explanation of $\alpha \vee \beta$? A related question: if γ is a preferred explanation of α and γ' entails γ, should γ' be considered a preferred explanation of α?

(b) Should changes on the background theory be allowed in other to explain an observation? and how much a change of the background theory affects explanations? For instance, suppose that γ is a preferred explanation of α with respect to Σ. Should γ be also a preferred explanation of α but now with respect to $\Sigma \cup \{\beta\}$?

There are many sources of motivating ideas for isolating the structural properties that will account for these basic questions. First of all, there is a vast literature on different areas of application of abduction: philosophy of science, linguistic, artificial intelligence, computer science, etc. All of them provide a large variety of examples where to look at for regularity patterns (see [5,20]). A second source of ideas is, of course, given by the structural properties of logical deduction (both classical and nonclassical). These structural properties has been studied (see [2,4]) in order to determine which of them could be considered valid for explanatory reasoning and how to modify those which are not valid in the context of abduction. For a comprehensive overview of abduction we refer the reader to [2,17]. The main idea used in this paper for isolating rules for explanatory reasoning will be explained in the following.

The examples given at the beginning of the introduction suggest that an important aspect of abduction is the set of conclusions to which the best explanation leads to. In other words, the consequences implied by the best explanation might be, in some cases, as relevant as the explanation itself. These considerations suggest that a measure of the "rationality" of an abductive method is given by the "rationality" of its "abductive consequences". More precisely, we view abduction as a binary relation between an observation and its preferred explanations. Following Flach's approach we work with a binary relation $\alpha \rhd \gamma$ between formulas which is read as saying *γ is a preferred explanation of α*. A *rationality postulate for explanatory reasoning* is a property of $\rhd$ saying that this relation is "well-behaved".

To each explanatory relation $\rhd$ we associate a consequence relation: given an observation α, we infer from α the common consequences of all preferred explanations of α. More formally, we define a consequence relation $\mid\!\sim_{ab}$ by

$$\alpha \mid\!\sim_{ab} \beta \quad \text{if } \Sigma \cup \{\gamma\} \vdash \beta \text{ for every } \gamma \text{ such that } \alpha \rhd \gamma. \tag{1}$$

We read $\alpha \mid\!\sim_{ab} \beta$ as "normally, if α is observed then β also should be present". In other words, β is a concomitant feature of every situation where α usually occurs.

The definition of $\mid\!\sim_{ab}$ is quite natural and, in fact, Levesque already suggested the idea of defining such consequence relation as a new deductive operation that would be useful when doing counterfactual experiments (see the concluding remarks of [13]). But the motivation to introduce this definition came from [15] where a consequence relation quite similar to $\mid\!\sim_{ab}$ was used to model abductive reasoning. Moreover, the results of [15] shows that $\mid\!\sim_{ab}$ has very nice formal properties. The key idea to isolate the postulates for explanatory reasoning is based in the interplay between $\rhd$ and $\mid\!\sim_{ab}$. We would like $\mid\!\sim_{ab}$ to be a bona fide consequence relation and for this end we have searched for postulates for $\rhd$ mainly guided by the well known rationality postulates for consequence relations studied by Kraus et al. [10], Makinson [16], Gärdenfors and Makinson [9] and many others.

We think that the use of the KLM methodology for isolating the postulates is not only an heuristic device but it also provides a fair enough justification for the postulates. The results of our analysis will give a formal justification for most of the postulates introduced by previous approaches and, in addition, it will shed new light on some aspects of abduction that we think have not been studied (this will be clarified in the following paragraphs).

In relation to (b) it is clear that these questions implicitly have the assumption that the background theory is also a parameter and thus that abduction is a ternary relation. This issue was addressed by Cialdea–Pirri and Aliseda who presented rules that allows some changes on Σ. However, they considered only changes that consists of adding new formulas to Σ. This restriction is quite natural, since more substantial changes (like contracting or revising Σ) are not a trivial matter as it is by now well known from the theory of belief revision developed by Gärdenfors and others [1,7]. In this paper the background theory will be fixed and therefore only formulas consistent with Σ can be explained. This can be considered a weakness since it has been argued that the more interesting observation are those which are not consistent with the theory ("surprising observations"). Boutilier and Becher [3] have presented a view of abduction based on the AGM theory for belief revision [1] by exploiting the idea that observations inconsistent with the background theory can be explained by revising the theory in order to make the observation either true or at least possible. At a first glance our approach seems to be incompatible with the belief revision approach because from this point of view Σ is considered a belief set and therefore as something defeasible. On the other hand, we will give Σ the role of a system description which is independent of the beliefs of the agent. The agent's believes are about which parts of the system are responsible for the observation but not about how the system is built. In other words, Σ represents the known laws of the world and base on them we explain an observation. [2] In spite of all this apparent differences, we will show in Section 4 that our approach also has an "epistemic" reading in the sense of belief revision.

(ii) Zadronzny, Cialdea–Pirri and Aliseda argued that abduction is a different form of reasoning and should not be reduced to reversed deduction. Flach's postulates reduces explanatory reasoning to reversed deduction (essentially because he did not include preference in his formalism. Nevertheless, his result goes in a direction similar to ours). The exact relationship between abduction and reversed deduction is however vague and, to our knowledge, has not being clarified in a formal way. We will say that an explanatory relation is *causal* if the following condition holds

$$\alpha \rhd \gamma \text{ iff } C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \gamma) \tag{2}$$

where $C_{ab}(\alpha) = \{\beta\colon \alpha \mid\!\sim_{ab} \beta\}$ and $\mid\!\sim_{ab}$ is defined as in (1) and $Cn(X)$ is the set of classical consequences of X (for X a set of formulas or a formula). We will argue in Section 3 that (2) can formally be regarded as saying that $\rhd$ and $\mid\!\sim_{ab}$ are dual objects and therefore that causal explanatory reasoning is nonmonotonic reasoning-in-reverse. We will see several examples of explanatory relations based on belief revision which are not causal (in our sense). These examples will show that the main feature of causal explanatory relations is

[2] A different but related problem is to repair Σ after some unexplainable fact is observed (or when the explanation are shown to be incorrect by other means). We think this problem is very close related with inductive reasoning and deserve a separated study.

that they are based on a non defeasible notion of explanation (as opposite to those notions based on belief).

(iii) As we have said one of the most distinct features of abduction is the emphasis it makes on preferred explanations rather than possible explanations. Most formalism we have mentioned include the notion of preference as an external requirement. Preference criteria for selecting the best explanation are regarded as qualitative properties (a sort of a simplicity criteria [3]) which are not reducible to logical ones. Moreover, in those formalism, the preference relation (for instance an order over formulas) is explicitly mentioned in the postulates that intend to capture the notion of "best" explanation. Cialdea and Pirri's approach tries to use preference criteria for selecting explanations based on logic but their results does not fully accomplish this goal since the preference relation has to be represented in a separated theory. In [19] we have shown that preference criteria are implicit in the logical properties of abduction and therefore they do not need to be explicitly included as part of the postulates. In other words, the structural properties of explanatory reasoning implicitly include an order encoding which are the preferred explanations. More formally, we have shown that (under some conditions) for every explanatory relation $\rhd$ there is an order relation $\prec$ such that $\alpha \rhd \gamma$ iff γ is a $\prec$-minimal explanation of α.

The paper is organized as follows. In Section 2 we will introduce and study the postulates for explanatory relations. In Section 3 we will show the tight relationship between our postulates and the rationality postulates for consequence relations in the KLM style. We will study causal explanatory relations and show that they are the formal counterpart of nonmonotonic consequence relations. In Section 4 we will see how our approach is viewed from the belief revision perspective. In Section 5 we will make precise comments about the work of Flach, Cialdea–Pirri, Aliseda and others. In Section 6 we will make some final remarks. Lists of the main postulates for consequence relations and explanatory relations used in the paper will be found in Appendixes A and B, respectively. A summary of the main results from Sections 2 and 3 will be given in Appendix C. The proofs will be given in Appendix D.

A preliminary version of this paper appeared as a technical report of LIFL 1997 (Lille, France) and part of it was presented at WOLLIC-97 (Brazil) and at NMR-98 (Italy).

2. Reasoning with explanations

The *background theory* denoted by Σ, will be a consistent set of formulas in a classical propositional language. We will use the following notation: $\alpha \vdash_\Sigma \beta$ when $\Sigma \cup \{\alpha\} \vdash \beta$. [4] We could have avoided the use of $\vdash_\Sigma$ and instead use a semantic entailment relation $\models$ satisfying the standard requirements (like compactness and the usual properties about $\vee$ and $\wedge$). This way the background theory would be taken for granted and the notion of explanation would be somewhat elliptical. But we have chosen to keep Σ for several reasons. First of all, because it is customary in most presentation of abduction to have a background theory. Secondly, because many examples are naturally presented with a

[3] Occam's razor: "*Entia praeter necessitatem non sunt multiplicanda*".

[4] Readers familiar with [15] should note that in that paper $\vdash_\Sigma$ denotes a different relation.

background theory that constrains the notion of explanation. And third, because by keeping Σ we leave open the question regarding the properties of abduction when the background theory is also considered a parameter.

We now introduce the notion of an explanation of a formula with respect to Σ.

Definition 2.1. For every formula α, the collection of explanations of α with respect to Σ is denoted by $Expla(\alpha)$ and is defined as follows:

$$Expla(\alpha) = \{\gamma\colon\ \gamma \nvdash_\Sigma \bot\ \ \&\ \gamma \vdash_\Sigma \alpha\}.$$

Notice that we have ruled out trivial explanations by asking that γ has to be consistent with Σ. We are interested in studying the relation "γ is a preferred explanation of α" which will be in most cases a proper subset of the relation "$\gamma \in Expla(\alpha)$". Our next definition capture some of the ideas mentioned in the introduction.

Definition 2.2. Let Σ be a background theory. An *explanatory relation* for Σ will be any binary relation $\rhd$ such that for every α and γ,

$$\alpha \rhd \gamma \Rightarrow \gamma \nvdash_\Sigma \bot \quad \text{and} \quad \gamma \vdash_\Sigma \alpha.$$

We read $\alpha \rhd \gamma$ as saying that γ is a preferred explanation (with respect to Σ) of α. The associated consequence relation is defined as follows

$$\alpha \mathrel{|\!\!\sim}_{ab} \beta \overset{\text{def}}{\Leftrightarrow} \gamma \vdash_\Sigma \beta \quad \text{for all } \gamma \text{ such that } \alpha \rhd \gamma.$$

We read $\alpha \mathrel{|\!\!\sim}_{ab} \beta$ as "normally, when α is observed then β should also be present". The collection of all abductive consequence of an observation $C_{ab}(\alpha)$ is defined as follows

$$C_{ab}(\alpha) = \{\beta\colon\ \alpha \mathrel{|\!\!\sim}_{ab} \beta\}.$$

In explanatory reasoning the input is an observation and the output is an explanation, that is the reason to write $\alpha \rhd \gamma$ with α as input and γ as output.

As we said in the introduction our initial and motivating idea was that $\mathrel{|\!\!\sim}_{ab}$ can be used heuristically to isolate the logical properties of explanatory relations. These properties will be called *postulates for explanatory reasoning*. We would like $\mathrel{|\!\!\sim}_{ab}$ to be a bona fide consequence relation and for this end we have searched for the postulates mainly guided by the well known KLM rationality postulates for consequence relations [10] (a list of the main postulates for consequence relations is given in Appendix A). The first thing we need is, of course, that $\mathrel{|\!\!\sim}_{ab}$ has to be reflexive, i.e., $\alpha \mathrel{|\!\!\sim}_{ab} \alpha$ for all α. This is obvious from the fact that when $\alpha \rhd \gamma$ then $\gamma \vdash_\Sigma \alpha$. Notice also that if $\alpha \vdash_\Sigma \beta$, then $\alpha \mathrel{|\!\!\sim}_{ab} \beta$. In particular, if $\alpha \vdash_\Sigma \bot$, then $\alpha \mathrel{|\!\!\sim}_{ab} \bot$.

A very natural assumption is to consider that explanatory relations are independent of the syntax. In our context this is expressed by the rules *Left Logical Equivalence* (**LLE**) and *Right Logical Equivalence* (**RLE**). Notice that these rules are somewhat stronger than the usual rules for consequence relations, since our notion of logical equivalence uses $\vdash_\Sigma$ instead of $\vdash$.

LLE If $\vdash_\Sigma \alpha \leftrightarrow \alpha'$ and $\alpha \rhd \gamma$, then $\alpha' \rhd \gamma$
RLE If $\vdash_\Sigma \gamma \leftrightarrow \gamma'$ and $\alpha \rhd \gamma$, then $\alpha \rhd \gamma'$

Next we introduce a postulate called *Explanatory Cautious Monotony* (**E-CM**), since it has the form of a monotonicity rule on the left.

E-CM If $\alpha \rhd \gamma$ and $\gamma \vdash_\Sigma \beta$, then $(\alpha \wedge \beta) \rhd \gamma$

This rule says that a preferred explanation γ of a simple observation α will be a preferred explanation of any observation more complex than α (like $\alpha \wedge \beta$) which is also entailed by γ. This seems quite natural because if we have decided that γ is a preferred explanation of α and we know further that γ implies β, then based on a larger set of observations (like $\alpha \wedge \beta$) it is reasonable to think that γ is a preferred explanation of $\alpha \wedge \beta$.

Now we will introduce the *Explanatory Cut rules*. These rules play an important role in our setting and, as we will see, there is a duality between monotony rules for consequence relations and cut rules for explanatory reasoning. Explanatory Cut rules relate the preferred explanations of an observation $\alpha \wedge \beta$ and the preferred explanations of α. If we have a complex observation (like $\alpha \wedge \beta$), then we might have an explanation for it which is not a preferred explanation for a simpler observation (like α). The observation of two facts (symptoms) together or simultaneously "forces" to select an explanation which might not be considered a preferred explanation when only one of the facts is observed. A Cut rule will say that, in certain cases, a preferred explanation of the more complex observation ($\alpha \wedge \beta$) might also be a preferred explanation of the simpler or incomplete observation (α). In other words, Cut rules allow to keep a preferred explanation even when the set of observations is not longer complete. One could get an idea of the usefulness of an Explanatory Cut rule by looking at a diagnosis process: if we know a fairly complete list of a patient's symptoms, then we might be able to decide which is the most likely illness that caused them. However, what if we know only few of the symptoms? An Explanatory Cut rule says that in certain cases this incomplete information suffices.

The first Cut rule we consider is the following

E-Cut If $(\alpha \wedge \beta) \rhd \gamma$, then $\beta \rhd \gamma$

This rule is quite strong as the following proposition shows.

Proposition 2.3. *Suppose $\rhd$ satisfies* **E-Cut**, *then $\mid\!\sim_{ab}$ is monotonic.*

Remark.

(i) It is easy to see that **E-Cut** is equivalent, under the presence of **E-CM**, to the following rule: If $\alpha \rhd \gamma$ and $\alpha \vdash_\Sigma \beta$, then $\beta \rhd \gamma$.

(ii) We consider **E-Cut** to be too strong to model the relation "γ is a preferred explanation of α". When γ is a *preferred* explanation of α, and α is an observation logically stronger than β (i.e., $\alpha \vdash \beta$), then the *preferred* explanations of β might not include γ, because we might need "less" to explain β than to explain α (an extreme case is when β is a consequence of Σ). We will present examples of natural explanatory relations which does not satisfy **E-Cut**. Among our cut rules, **E-Cut** is essentially the only Cut rule we have seen in the literature.

We will consider in this paper two others Cut rules: Explanatory Cautious Cut (**E-C-Cut**) and Explanatory Rational Cut (**E-R-Cut**).

E-C-Cut If $(\alpha \wedge \beta) \rhd \gamma$ and $[\alpha \rhd \delta \Rightarrow \delta \vdash_\Sigma \beta]$ for all δ, then $\alpha \rhd \gamma$
E-R-Cut If $(\alpha \wedge \beta) \rhd \gamma$ and there is δ such that $\delta \vdash_\Sigma \beta$ and $\alpha \rhd \delta$, then $\alpha \rhd \gamma$

Remark.

(i) Cut rules are the key fact for encoding preference criteria. Suppose $(\alpha \wedge \beta) \rhd \gamma$ and $\alpha \not\rhd \gamma$. This can be interpreted as saying that some part (β) of the observation $\alpha \wedge \beta$ is more important than the other part (α) and therefore it cannot be ignored when selecting the preferred explanations of the complete observation $\alpha \wedge \beta$. This will be clarified in the examples (see Section 2.1).

(ii) The meaning **E-C-Cut** is more easily grasp by analyzing its contrapositive: suppose $(\alpha \wedge \beta) \rhd \gamma$ and $\alpha \not\rhd \gamma$, then there exists δ such that $\alpha \rhd \delta$ and $\delta \nvdash_\Sigma \beta$. It says, in particular, that if we are able to find a good explanation for $\alpha \wedge \beta$, then we should also be able to find a good explanation for α (but maybe a different one). **E-R-Cut** can be stated in an equivalent form as follows: if γ is a good explanation of $\alpha \wedge \beta$ but it is not a good explanation of α then any good explanation of α is consistent with $\neg\beta$.

(iii) In [19] we show that **E-R-Cut** implies that preferred explanations (i.e., those formulas γ such that $\alpha \rhd \gamma$ for some α) are linearly pre-order. Moreover, when the underlying language is finite, **E-R-Cut** turns out to be equivalent to assigning a natural number to each formula and thus the preferred explanation of α are those explanations of α with minimal value.

In general $\rhd$ is not reflexive, because a formula might not be a *preferred* explanation of itself (this was already noticed in [4,5]), however there is a version of reflexivity that holds in most cases.

E-Reflexivity If $\alpha \rhd \gamma$, then $\gamma \rhd \gamma$

Suppose that **E-CM** and **E-C-Cut** hold. Let $\alpha \rhd \gamma$, then by **E-CM** we have $(\gamma \wedge \alpha) \rhd \gamma$. It is easy to check that the hypothesis of **E-C-Cut** are satisfied and hence $\gamma \rhd \gamma$. So we have shown the following

Proposition 2.4. *Let $\rhd$ an explanatory relation satisfying* **E-CM** *and* **E-C-Cut**. *Then* **E-Reflexivity** *holds.*

The following result shows that the postulates for explanatory relations considered so far are the counterpart of cumulative consequence relations, i.e., relations satisfying the following rules:

REF	(Reflexivity)	$\alpha \mid\!\sim \alpha$
LLE	(Left Logical Equivalence)	$\alpha \mid\!\sim \beta \;\&\; \vdash \alpha \leftrightarrow \gamma \Rightarrow \gamma \mid\!\sim \beta$
RW	(Right Weakening)	$\alpha \mid\!\sim \beta \;\&\; \vdash \beta \to \gamma \Rightarrow \alpha \mid\!\sim \gamma$
CUT		$\alpha \wedge \beta \mid\!\sim \gamma \;\&\; \alpha \mid\!\sim \beta \Rightarrow \alpha \mid\!\sim \gamma$
CM	(Cautious Monotony)	$\alpha \mid\!\sim \beta \;\&\; \alpha \mid\!\sim \gamma \Rightarrow \alpha \wedge \gamma \mid\!\sim \beta$

Theorem 2.5. *Suppose $\rhd$ satisfies* **LLE**, **E-CM** *and* **E-C-Cut**, *then $\mid\!\sim_{ab}$ is cumulative.*

Now will address the problem of how explanatory relations treat disjunctions. We will start by analyzing the right side. Consider the following postulates:

E-RW If $\alpha \rhd \gamma$ and $\alpha \rhd \delta$, then $\alpha \rhd (\gamma \vee \delta)$
ROR If $\alpha \rhd (\gamma \vee \rho)$, then $\alpha \rhd \gamma$ or $\alpha \rhd \rho$
E-Disj If $\alpha \rhd (\gamma \vee \rho)$ and $\gamma \nvdash_\Sigma \bot$, then $\alpha \rhd \gamma$
RA If $\alpha \rhd \gamma$, $\gamma' \vdash_\Sigma \gamma$ and $\gamma' \nvdash_\Sigma \bot$, then $\alpha \rhd \gamma'$

Remark.

(i) In [13] it was argued that if α has more than one preferred explanation, then the disjunction of all of them is the explanation that fully and nontrivially accounts for α. The consequence relation $\mid\!\sim_{ab}$ is capturing this intuition, since to compute the abductive consequences of α is irrelevant whether the collection of preferred explanations of α is closed under disjunctions. These considerations suggest **E-RW**. This postulate will be called *Explanatory Right Weakening*. It is the only rule that allows to weakening a preferred explanation. In Section 3 we will present a natural family of explanatory relations satisfying **E-RW**.

(ii) Postulate **ROR** and **E-Disj** are called *Right Or* and *Explanatory Disjunction*, respectively. Notice that **ROR** is weaker than **E-Disj**. We will show below that **RA** is equivalent to **E-Disj** plus **RLE**. Postulate **RA** will be called *Right And* since it gives some amount of monotony on the right. A similar postulate has been considered by Flach [5]. **RA** says that any explanation more "complete" (i.e., logically stronger) than a preferred explanation of α is also a preferred explanation of α. In Section 4 we will show that explanatory relations satisfying **RA** are based on a nondefeasible notion of explanation.

Proposition 2.6. *Let $\rhd$ be an explanatory relation.*

(i) *If $\rhd$ satisfies* **RA**, *then it satisfies* **RLE** *and* **ROR**.

(ii) *Suppose $\rhd$ satisfies* **RA**. *If $\alpha \rhd \gamma$ and $\gamma \rhd \delta$, then $\alpha \rhd \delta$. In other words, $\rhd$ is transitive.*

(iii) *Suppose $\rhd$ satisfies* **E-CM** *and* **RA**. *If $\alpha \rhd \gamma$ and $\gamma \nvdash_\Sigma \neg\beta$, then there is $\gamma' \vdash_\Sigma \gamma$ such that*

$$\alpha \rhd \gamma',\ \gamma' \vdash_\Sigma \beta \quad \textit{and} \quad (\alpha \wedge \beta) \rhd \gamma'.$$

(iv) **E-Disj** *together with* **RLE** *is equivalent to* **RA**.

(v) *Suppose $\rhd$ satisfies* **E-CM**, **LLE** *and* **RA**. *Then*

$$\bigl\{\gamma\colon (\alpha \vee \beta) \rhd \gamma\bigr\} \subseteq \{\gamma\colon \alpha \rhd \gamma\} \cup \{\gamma\colon \beta \rhd \gamma\} \cup \bigl\{\gamma\colon \vdash_\Sigma \gamma \leftrightarrow (\gamma_1 \vee \gamma_2),\ \alpha \rhd \gamma_1, \beta \rhd \gamma_2\bigr\}.$$

Definition 2.7. An explanatory relation is said to be *E-preferential* if satisfies **LLE**, **E-CM**, **E-C-Cut** and **RA**.

The next result says that E-preferential explanatory relations captures our initial motivation for introducing the postulates. Recall that a consequence relation $\mid\!\sim$ is preferential if in addition to cumulative rules $\mid\!\sim_{ab}$ satisfies the rule Or: for any formulas α, β and γ if $\alpha \mid\!\sim_{ab} \gamma$ and $\beta \mid\!\sim_{ab} \gamma$ then $\alpha \vee \beta \mid\!\sim_{ab} \gamma$.

Theorem 2.8. *If $\rhd$ is an E-preferential explanatory relation, then $\mid\!\sim_{ab}$ is preferential.*

Remark. It is interesting to observe the analogy between Proposition 2.6(v) and the fact that for a preferential consequence relation $\mid\!\sim$, $C(\alpha) \cap C(\beta) \subseteq C(\alpha \vee \beta)$ (where $C(\alpha)$ denotes the set $\{\beta\colon \alpha \mid\!\sim \beta\}$). In other words, the sets $\{\gamma\colon \alpha \rhd \gamma\}$ and $C(\alpha)$ seem to play dual roles.

We will continue using the properties of $\mid\!\sim_{ab}$ as a guideline for isolating rationality postulates for abduction. We will consider next the following postulates:

WDR	(*Weak Disjunctive Rationality*)	$C(\alpha \vee \beta) \subseteq Cn(C(\alpha) \cup C(\beta))$
DR	(*Disjunctive Rationality*)	if $\alpha \vee \beta \mid\!\sim \rho$ then either $\alpha \mid\!\sim \rho$ or $\beta \mid\!\sim \rho$
RM	(*Rational Monotony*)	if $\alpha \mid\!\sim \rho$ and $\alpha \not\mid\!\sim \neg\beta$, then $\alpha \wedge \beta \mid\!\sim \rho$

These rules has been studied both from a semantics point of view [6,12] and a syntactical point of view [16]. The new postulates for $\rhd$ will be related to properties satisfied by the preferred explanations of a disjunctive formula. Which is not surprising, since **WDR**, **DR** and **RM** impose constrains to the set of consequences of a disjunctive formula.

We will use two postulates for the left side:

LOR	If $\alpha \rhd \gamma$ and $\beta \rhd \gamma$, then $(\alpha \vee \beta) \rhd \gamma$
E-DR	If $\alpha \rhd \gamma$ and $\beta \rhd \delta$, then $(\alpha \vee \beta) \rhd \gamma$ or $(\alpha \vee \beta) \rhd \delta$

Remark.

(i) **LOR** is called *Left Or*. The intuition behind **LOR** is the following. Suppose that when we observe either α or β (no matter which one) we are willing to accept that γ is a very likely explanation for both of them. Now we are told that one of them is observed (but maybe it is not known which one). Is it rational to conclude that γ is still a very likely explanation of that observation (i.e., a very likely explanation of $\alpha \vee \beta$)? **LOR** implies that the answer is yes. It is interesting to notice that **LOR** was considered by Flach and Aliseda as a principle for confirmatory induction rather than for explanatory inference.

(ii) We will show below that **LOR** corresponds to **WDR**. Freund [6] proved that, in the case of finite languages, a preferential relation satisfies **WDR** iff it can be represented by an injective preferential model.

(iii) It is easy to check that **DR** is equivalent to saying that $C(\alpha \vee \beta) \subseteq C(\alpha) \cup C(\beta)$ for every α and β. Hence, **DR** is stronger than **WDR**. We will show that the corresponding postulate for explanatory relations is **E-DR** and thus we have called it *Explanatory Disjunctive Rationality*.

Theorem 2.9. *Suppose the language is finite and let $\rhd$ be an E-preferential explanatory relation that satisfies* **LOR**. *Then $\mid\!\sim_{ab}$ is preferential and satisfies* **WDR**.

Remark. We do not know if Theorem 2.9 holds when the language is infinite.

Proposition 2.10. *Let $\rhd$ be an explanatory relation satisfying* **E-DR**. *Then $\rhd$ satisfies* **LOR** *and $\mid\!\sim_{ab}$ satisfies* **DR**.

As a corollary of Theorems 2.8 and 2.10 we have

Theorem 2.11. *Let $\rhd$ be an E-preferential explanatory relation that satisfies* **E-DR**. *Then $\mid\!\sim_{ab}$ is preferential and satisfies* **DR**.

A relation $\mid\!\sim$ is called *Rational* if it is preferential and satisfies Rational Monotony (**RM**). The corresponding postulate for abduction is the cut rule we have called **E-R-Cut**. We recall it: If $(\alpha \wedge \beta) \rhd \gamma$ and there is δ such that $\delta \vdash_{\Sigma} \beta$ and $\alpha \rhd \delta$, then $\alpha \rhd \gamma$.

Theorem 2.12. *Let $\rhd$ be an E-preferential explanatory relation that satisfies* **E-R-Cut**. *Then $\mid\!\sim_{ab}$ is rational.*

We will see next that **E-R-Cut** gives a fine structure to the set $\{\gamma\colon (\alpha \vee \beta) \rhd \gamma\}$.

Proposition 2.13. *Suppose $\rhd$ is an E-preferential explanatory relation that satisfies* **E-R-Cut**. *Then for every α and β one of the following holds*:

(a) $\{\gamma\colon (\alpha \vee \beta) \rhd \gamma\} = \{\gamma\colon \alpha \rhd \gamma\}$,

(b) $\{\gamma\colon (\alpha \vee \beta) \rhd \gamma\} = \{\gamma\colon \beta \rhd \gamma\}$,

(c) $\{\gamma\colon \alpha \rhd \gamma\} \cup \{\gamma\colon \beta \rhd \gamma\} \subseteq \{\gamma\colon (\alpha \vee \beta) \rhd \gamma\} \subseteq$
$\{\gamma\colon \alpha \rhd \gamma\} \cup \{\gamma\colon \beta \rhd \gamma\} \cup \{\gamma\colon \vdash_{\Sigma} \gamma \leftrightarrow (\delta \vee \rho)\ \&\ \alpha \rhd \delta\ \&\ \beta \rhd \rho\}$.

Remark. The second $\subseteq$ in (c) above could be an equality if $\rhd$ satisfies **E-RW**. In this case, Proposition 2.13 is the analogous of the following well known fact about rational relations (which was found first in the context of belief revision [7,8]): If $\mid\!\sim$ is rational then for every α and β one of the following holds:

(a) $C(\alpha \vee \beta) = C(\alpha)$,

(b) $C(\alpha \vee \beta) = C(\beta)$,

(c) $C(\alpha \vee \beta) = C(\alpha) \cap C(\beta)$.

The proof of Proposition 2.13 follows closely the proof of this fact about $\mid\!\sim_{ab}$.

It is well known that any rational relation satisfies **DR** [16]. We will show next the corresponding result for **E-DR** (it will be used later in the paper).

Proposition 2.14. *Suppose $\rhd$ is E-preferential and satisfies* **E-R-Cut**. *Then it also satisfies* **E-DR**.

On the light of the previous results we will complete Definition 2.7 as follows

Definition 2.15. Let Σ be a background theory and $\rhd$ be an explanatory relation. We say that $\rhd$ is **E-cumulative** if it satisfies **E-CM**, **E-C-Cut** and **LLE**. $\rhd$ is **E-preferential** if it is E-cumulative and in addition satisfies **RA**. $\rhd$ is **E-rational** if it is E-preferential and in addition satisfies **E-R-Cut**.

We are about to finish the presentation of the postulates for explanatory reasoning. There is however one natural question that we have not considered yet: When an observation has a preferred explanation? The following postulate, that we call *Explanatory Consistency Preservation*, says that α has a preferred explanation iff it is consistent with Σ. Our last results are somewhat technical but they will be needed in the sequel.

E-Con$_\Sigma$: $\nvdash_\Sigma \neg\alpha$ iff there is γ such that $\alpha \rhd \gamma$.

The corresponding postulate for consequence relations will be called *Consistency Preservation* (with respect to Σ).

Con$_\Sigma$: For every formula α, (i) $\alpha \mid\!\sim \bot$ iff $\vdash_\Sigma \neg\alpha$ and (ii) for every $\sigma \in \Sigma$, $\alpha \mid\!\sim \sigma$.

Part (ii) in **Con$_\Sigma$** was included since it necessarily holds for $\mid\!\sim_{ab}$. The following observation is obvious.

Proposition 2.16. *Let $\rhd$ be an explanatory relation satisfying* **E-Con$_\Sigma$**, *then $\mid\!\sim_{ab}$ satisfies* **Con$_\Sigma$**.

Under **E-Con$_\Sigma$**, **E-R-Cut** is stronger than **E-C-Cut**. More precisely we have the following

Proposition 2.17. *Any explanatory relation satisfying* **E-Con$_\Sigma$** *and* **E-R-Cut** *satisfies* **E-C-Cut**.

As a corollary of Propositions 2.14 and 2.17 we have the following result:

Proposition 2.18. *Suppose that $\rhd$ satisfies* **LLE**, **E-CM**, **RA**, **E-R-Cut** *and* **E-Con$_\Sigma$**. *Then it also satisfies* **E-DR**.

As a corollary of Theorem 2.12 and Propositions 2.16, 2.17 we have the following result:

Proposition 2.19. *Let $\rhd$ be an explanatory relation that satisfies* **LLE**, **E-CM**, **E-R-Cut**, **E-Con$_\Sigma$**, *and* **RA**. *Then $\mid\!\sim_{ab}$ is rational and satisfies* **Con$_\Sigma$**.

Proposition 2.20. *Suppose $\rhd$ satisfies* **E-Cut** *and* **E-Con$_\Sigma$**, *then $\mid\!\sim_{ab} = \vdash_\Sigma$.*

2.1. Two examples

We will present examples of E-preferential and E-rational explanatory relations. Both examples are based on preferential models. Preferential models are the main tool for representing and studying nonmonotonic consequence relations (see [10] and references therein). Given an order of the models of Σ we define a notion of preferred explanation. The intuition is that to explain an observation we only look at the closest worlds where the observation holds. We will use the following notation: $mod(S)$ denotes the set of models of S, where S is a set of formulas (it could be a single formula). The general idea is the following: Given a preference relation $\prec$ on $mod(\Sigma)$ and a formula α we define its minimal models as usual:

$$min(\alpha) = \bigl\{N\colon\ N \models \Sigma \cup \{\alpha\}\ \&\ M \not\models \alpha \text{ for all } M \prec N\bigr\}.$$

Now we define an explanatory relation $\rhd$ as follows:

$$\alpha \rhd \gamma \overset{\text{def}}{\Leftrightarrow} mod\bigl(\Sigma \cup \{\gamma\}\bigr) \subseteq min(\alpha)$$

for any pair of consistent (with Σ) formulas α and γ. [5]

It is not an accident that we use preferential models. In fact explanatory relations defined this way are quite universal in the sense that many explanatory relations are of that form (this will be addressed in Section 3).

We could have presented the examples just as a formal manipulation of symbols, but instead we choose to provide a context where to interpret the symbols. This kind of interpretations (that makes the reading more enjoyable) have a drawback: important aspects of the context are not included into the formalism used to model it; so one get the impression that the formalism is an over simplification of the problems under consideration. Our examples mainly pretend to illustrate some of the concepts we have introduced.

Example 1. Consider the following scenario. A message consisting of a finite sequence of 0 and 1 is sent by either one of two independent senders A or B. Messages sent by A always start with 0 and messages sent by B always start with 1. Sometimes only a portion of the message is received and thus it is necessary to recover the lost part. The person in charge of recovering messages, after many years of persistent work, has developed a quite simple preference criterion for guessing the correct message. He has observed that normally both A and B send messages starting with a constant sequence and moreover the sequence has even length. Since the senders are independent of each other he has not preference about who sends the message. To make the example manageable we will assume that all messages have length 4. We will analyze later in the paper a similar example allowing messages of any length.

The preference criterion can then be represented as follows:

$$\begin{array}{ccc}
\{0100, 0101, 0110, 0111, 0001\} & \qquad & \{1000, 1001, 1010, 1011, 1110\} \\
| & & | \\
| & & | \\
\{0000, 0010, 0011\} & & \{1111, 1100, 1101\}
\end{array}$$

[5] When the language is infinite or $\prec$ is not transitive it is necessary to require the so called *smoothness condition*: for all formula α and all $N \models \Sigma \cup \{\alpha\}$ which is not in $min(\alpha)$, there is $M \models \Sigma \cup \{\alpha\}$ such that $M \prec N$ and $M \in min(\alpha)$. This condition obviously holds if $\prec$ is a well-founded pre-order, which will be the type of relations used in this paper.

Where the messages at the bottom are more preferred than those at the top, but there is no relation between a message starting with 0 and a message starting with 1.

Let the letters a, b, c and d represent, in that order, the four digits of a message. The language $\mathcal{L}$ is the propositional language in the variables a, b, c and d and Σ is the empty set (any message can be either sent or received and there is no logical connection between the digits of a given message). Every message is a valuation of $\mathcal{L}$ and therefore the preference relation described above is a partial order over the collection of all interpretations of the language. This partial order will be denoted by $\prec$. Notice that all valuations at the bottom (or top) are mutually incomparable. Given a formula α we define its minimal models $min(\alpha)$ as we said at the beginning of this section. We interpret $min(\alpha)$ as containing those messages encoded by α that have the most preferred features. Thus our definition says that γ is a preferred explanation for α if every message encoded by γ is one of the preferred messages encoded by α. This is not quite the same as saying that every *preferred* message encoded by γ is also one of the preferred messages encoded by α. The last statement holds if we ask that $min(\gamma) \subseteq min(\alpha)$. This alternative will be considered later.

It is easy to show that $\alpha \mid\!\sim_{ab} \beta$ iff $N \models \beta$ for all $N \in min(\alpha)$. This can be stated equivalently as $mod(C_{ab}(\alpha)) = min(\alpha)$. Readers familiar with the theory of nonmonotonic consequence relations will realize the motivation for our definition. We will make this connection clear in the forthcoming sections.

It is not difficult to show that $\rhd$ is a E-preferential explanatory relation. We will not proved this now since it is a consequence of a general result that will be shown later (see Section 3.2). We will compute some preferred explanations.

Suppose that the portion of the message we were able to get is expressed by the formula d (i.e., we only know that the fourth digit is 1). Then it is easy to check that the most likely sent messages are 0011, 1101 or 1111. Thus the preferred explanation of d are $\neg a \wedge \neg b \wedge c \wedge d$, $a \wedge b \wedge \neg c \wedge d$, $a \wedge b \wedge c \wedge d$ and the disjunction of them. In particular, $\rhd$ is not reflexive, for instance $d \not\rhd d$. Notice that $d \mid\!\sim_{ab} (\neg a \wedge \neg b \,\vee\, a \wedge b)$, which reflects the agent's preferences.

Let us suppose that in addition we know that the second digit was 0. Now the observation is encoded by $\neg b \wedge d$. In this case the most likely sent messages are 0011, 1001 and 1011. The formulas encoding these messages together with their disjunction are all the preferred explanation of $\neg b \wedge d$. Notice that **E-R-Cut** fails. In fact, 1001 is a preferred explanation of $d \wedge \neg b$ which is not a preferred explanation of d but there is a preferred explanation of d (namely 0011) that implies $\neg b$.

We have already suggested that there are other natural alternatives to define $\rhd$ based on a preferential model. For example, requiring that $min(\gamma) \subseteq min(\alpha)$ instead of $mod(\gamma) \subseteq min(\alpha)$. The main difference of this alternative definition with respect to the one given above is that the former is reflexive and fails to satisfy **RA** but the later is not. This will be treated in Section 4.

Example 2. Leonidas, an old taxi driver, retired two month ago after 50 years of work. He lent his car to Julio, a nephew of him. Every time Leonidas has an opportunity he enjoyed himself by guessing which streets his nephew has driven his car by. Leonidas just needs to ask a couple of questions and then he is able to tell very precisely the exact route Julio took. He uses to say, making fun of Julio, "*my car is more like a metro train*

$$
\begin{array}{cccccccc}
P & \leftarrow & (2,4) & \leftarrow & (3,4) & \leftarrow & H & \text{4th St.} \\
\downarrow & & \uparrow & & \downarrow & & \uparrow & \\
(1,3) & \rightarrow & (2,3) & \rightarrow & (3,3) & \rightarrow & (4,3) & \text{3rd St.} \\
\downarrow & & \uparrow & & \downarrow & & \uparrow & \\
(1,2) & \leftarrow & (2,2) & \leftarrow & K & \leftarrow & (4,2) & \text{2nd St.} \\
\downarrow & & \uparrow & & \downarrow & & \uparrow & \\
C & \rightarrow & (2,1) & \rightarrow & (3,1) & \rightarrow & (4,1) & \text{1th St.} \\
 & & & & & & & \\
\text{1th Ave.} & & \text{2nd Av.} & & \text{3rd Av.} & & \text{4th Av.} &
\end{array}
$$

Fig. 1. *Chacaito* station is at C, *La Hoyada* station is at H, *Café Kawi* is at K and *Cine Paraíso* is at P.

that needs no driver and you are in the car not really to drive it but only to collect the fare". Once he got into a big trouble by trying to impress his nephew with his divining skills. He could not help himself and approached a young couple that just got off the car. Very politely he addressed them with his usual questions: "*Where did you get in*?", "*Did you pass by* Café Kawi?", "*Did you pass by* Cine Paraíso?" The young couple got into a awful argument. The outburst, Leonidas and Julio thought, had nothing to do with the questions they asked. The young man said "*we did pass by Café Kawi but not by the movie theater*" and she replied, "*as usual, you were absent mind, thinking about god knows what*! *We did not pass by the Café but we did pass by the theater*". That day Julio made his uncle swear that he will never again bother his customers with such nagging questions. The old taxi driver slowly walked away, then turned his head and smiling said to Julio "*You did pass by the movie theater, anyway*". The reason for Leonidas' success in guessing the routes was that he has given Julio very precise indications about which were the best routes for avoiding traffic and finding good customers. He said to Julio: "*Always try to pass by either one of the two metro station* Chacaito *or* La Hoyada. *In case this is not possible, then try to pass by either* Café Kawi *or* Cine Paraíso. *If neither of these two alternatives are possible, do whatever you feel like*". Julio always follows Leonidas' advice to the letter.

The street map of the area covered by Leonidas' car is indicated in Fig. 1.

To model this example we introduce one propositional variable $z_{i,j}$ for each one of the 16 corners in the map. It is also convenient to add another 32 new variables to denote the starting and ending points. Let $s_{i,j}$ denote that the starting point was at (i, j) and similarly $e_{i,j}$ for the ending point. The intended models (i.e., taxi rides) will be paths through this map. We will only consider paths satisfying the following constrains:

(i) a path has a unique starting and ending point,

(ii) a path should not intersect itself, and

(iii) a path can have only one connected component. Then Σ is the theory of all these intended models.

Leonidas's preferences are given by a three level preferential model.

$$L_0 = mod\big(\Sigma \cup \{z_{1,1} \vee z_{4,4}\}\big),$$
$$L_1 = mod\big(\Sigma \cup \{\neg z_{1,1} \wedge \neg z_{4,4},\ z_{3,2} \vee z_{1,4}\}\big),$$
$$L_2 = mod(\Sigma) \setminus (L_0 \cup L_1).$$

This gives a total pre-order (i.e., a transitive and reflexive relation) of $mod(\Sigma)$. The explanatory relation $\rhd$ is defined as explained at the beginning of this section. So $min(\alpha)$ consists of those models of $\Sigma \cup \{\alpha\}$ which are minimal with respect to the pre-order defined above. A general result, which will be proved later, guarantees that $\rhd$ is E-rational, satisfies **E-RW** and, since $\Sigma = L_0 \cup L_1 \cup L_2$ then $\rhd$ also satisfies **E-Con**$_{\boldsymbol{\Sigma}}$ (see Section 3.2).

Let us suppose that the couple got in the car at $(3,4)$ and off the car at $(2,2)$. Let α be $(z_{3,2} \wedge \neg z_{1,4} \vee \neg z_{3,2} \wedge z_{1,4}) \wedge s_{3,4} \wedge e_{2,2}$. Notice that Leonidas had the information encoded by α. Since α has models in L_0, then a preferred explanation of α must be a formula γ such that $mod(\gamma) \subseteq L_0$ and $\gamma \vdash_\Sigma \alpha$. It is clear that any path starting at $(3,4)$ and ending at $(2,2)$ cannot pass by H. Hence any preferred explanation of α necessarily is a path passing by C. From this it is easy to check that there is only one solution and it includes P. Notice that there are several formulas describing this unique solution. For instance, $s_{3,4} \wedge z_{1,3} \wedge z_{2,1} \wedge e_{2,2}$. We do not need to mention all corners in this path. Some of them will be forced to be in the path by the rules of Σ. Observe that the preferred explanations of α are exactly the preferred explanations of $\neg k \wedge p \wedge s_{3,4} \wedge e_{2,2}$ (here recall Proposition 2.13).

Let β be the following "observation" $s_{2,1} \wedge z_{2,2} \wedge z_{2,3} \wedge e_{2,4} \wedge \neg k \wedge \neg z_{3,4}$. So β encodes partial information about a ride that started at $(2,1)$ and ended at $(2,4)$, passed by $(2,2)$, $(2,3)$ and did not pass neither by Café Kawi nor by $(3,4)$. Any path satisfying β starts at $(2,1)$, then it cannot pass by C and since it does not pass by $(3,4)$ then it cannot pass by H. In fact, we have that $\beta \vdash_\Sigma \neg z_{1,4} \wedge \neg z_{3,2} \wedge \neg z_{1,1} \wedge \neg z_{4,4}$. This says that all models of β belong to L_2. Therefore the preferred explanations of β are formulas all whose models must be in L_2. What if we do not know the starting point? For instance, let α be $z_{2,2} \wedge z_{2,3} \wedge e_{2,4} \wedge \neg k \wedge \neg z_{3,4}$. This observation is a weaker than β and moreover α has models in L_0 (for instance a path starting at C, then it goes to $(2,1)$, then goes through 2nd Ave. and finally stops at $(2,4)$). Hence none of the preferred explanations of β is a preferred explanation of α. This example shows that some parts of an observation are more important (because they are more relevant) than others and therefore cut rules must be constrained.

Let now β' be the following formula: $s_{2,1} \wedge z_{2,3} \wedge e_{2,4} \wedge \neg k \wedge \neg z_{3,4}$. We claim that the preferred explanations of β' are exactly the preferred explanations of β. In fact, it is easy to check that there are preferred explanations of β' that implies $z_{2,2}$. Then by **E-R-Cut** we conclude that any preferred explanation of β is also a preferred explanation of β'. This says that in this case $z_{2,2}$ is irrelevant and therefore can be ignored.

To relate the meaning of **E-R-Cut** with the ranked model that defines the explanatory relation, let us suppose that $(\alpha \wedge \beta) \rhd \gamma$. The constrain in **E-R-Cut** says that there must exist δ such that $\alpha \rhd \delta$ and $\delta \vdash_\Sigma \beta$. This implies that $min(\alpha)$ are at the same level as $min(\alpha \wedge \beta)$, therefore γ remains a preferred explanation for α.

3. Explaining our reasoning

In the previous section we have shown that each explanatory relation has associated a consequence relation which reflects many properties of the explanatory relation. The intuition was: *if you tell me how to explain an observation, then I will tell you which are its usual or normal consequences*. In this section we will address the converse of the previous statement: *If you know which are the normal consequences of an observation, can you explain it*? In this setting there are two obvious thing one has to remark. The first one is that we are viewing the process of getting conclusions out of an observation and the process of explaining it as dual processes. But then it is natural to ask: are these two processes one the inverse of the other? To answer this question we will introduce a notion of *causal explanatory relation* and show that it corresponds to explanatory mechanisms that can be formally regarded as performing reversed nonmonotonic deduction.

The normal consequences of an observation will be given by a consequence relation $\mid\!\sim$. We will assume that every such $\mid\!\sim$ is reflexive, i.e., $\alpha \mid\!\sim \alpha$ for all α. The first thing we must answer is under which conditions $\mid\!\sim$ is of the form $\mid\!\sim_{ab}$. It is obvious from the definition of $\mid\!\sim_{ab}$ that the question is then when the following holds:

$$C(\alpha) = \bigcap \bigl\{Cn(\Sigma \cup \{\gamma\}) \colon C(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})\bigr\}. \tag{3}$$

We formally introduce this condition in the following definition.

Definition 3.1. A consequence relation $\mid\!\sim$ is said to be *adequate with respect to* Σ if (3) holds for every formula α.

If $\rhd$ is an explanatory relation then, from the definition of $\mid\!\sim_{ab}$, it is clear that $\mid\!\sim_{ab}$ is adequate with respect to Σ. The classical entailment relation $\vdash$ is adequate with respect to $\{\top\}$ and $\vdash_\Sigma$ is adequate with respect to Σ. If there is no danger of confusion we will just say *adequate* instead of *adequate with respect to* Σ.

Given an adequate with respect to Σ consequence relation $\mid\!\sim$ it is clear that $\alpha \mid\!\sim \sigma$ for all $\sigma \in \Sigma$. Moreover, if $\alpha \not\mid\!\sim \bot$, then there must exist γ consistent with Σ such that $\gamma \vdash_\Sigma \alpha$. In particular, if $\alpha \not\mid\!\sim \bot$ then α is consistent with Σ. Hence $\mid\!\sim$ almost satisfies $\mathbf{Con}_\Sigma$ except that it might happen that $\alpha \mid\!\sim \bot$ for some α consistent with Σ. Also observe that an adequate consequence relation satisfies the following form of supraclassicality: if $\alpha \vdash_\Sigma \beta$, then $\alpha \mid\!\sim \beta$.

The notion of an adequate consequence relation is relevant only if the language is infinite. In fact, for a finite language, it is not hard to show that every consequence relation satisfying the following mild conditions is adequate: (i) $C(\alpha) = Cn(C(\alpha))$ and (ii) $\alpha \mid\!\sim \sigma$ for all α and all $\sigma \in \Sigma$. However, for infinite languages there are even rational relations satisfying $\mathbf{Con}_\Sigma$ which are not adequate (see Example 5 in Section 3.2).

It is clear from (3) what should be the definition of the explanatory relation associated with a consequence relation.

Definition 3.2. Let $\mid\!\sim$ be a consequence relation $\mid\!\sim$. We associate with $\mid\!\sim$ a binary relation $\tilde{\rhd}$ as follows:

$$\alpha \mathrel{\tilde{\rhd}} \gamma \overset{\text{def}}{\Leftrightarrow} \gamma \nvdash_\Sigma \bot \ \ \& \ C(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\}). \tag{4}$$

Notice that $\tilde{\rhd}$ is indeed an explanatory relation (using that $\mid\!\sim$ is reflexive). We have put a tilde above the symbol $\rhd$ to remind the reader that this explanatory relation is defined using a consequence relation $\mid\!\sim$. Suppose that $\mid\!\sim$ satisfies the following form of supraclassicality: if $\alpha \vdash_\Sigma \beta$, then $\alpha \mid\!\sim \beta$. Then it is clear that if $\alpha \mathrel{\tilde{\rhd}} \gamma$ then $\gamma \mid\!\sim \alpha$. However, in general $\gamma \mid\!\sim \alpha$ does not imply $\alpha \mathrel{\tilde{\rhd}} \gamma$ as we will see in the examples. This suggests an alternative definition which will be treated in Section 4.

The following result is easy to show.

Proposition 3.3. *Every adequate consequence relation is of the form* $\mid\!\sim_{ab}$.

The next theorem shows the correspondence between the postulates satisfied by $\mid\!\sim$ and those satisfied by $\tilde{\rhd}$.

Theorem 3.4. *Let* $\mid\!\sim$ *be an adequate consequence relation, then*:

(1) $\tilde{\rhd}$ *satisfies* **RA**, **E-RW** *and* **RLE**.
(2) *If* $\mid\!\sim$ *satisfies* **LLE**, *then* $\tilde{\rhd}$ *satisfies* **LLE**.
(3) *If* $\mid\!\sim$ *satisfies* $\mathbf{Con}_{\boldsymbol{\Sigma}}$, *then* $\tilde{\rhd}$ *satisfies* $\mathbf{E\text{-}Con}_{\boldsymbol{\Sigma}}$.
(4) *If* $\mid\!\sim$ *satisfies* **CM**, *then* $\tilde{\rhd}$ *satisfies* **E-C-Cut**.
(5) *If* $\mid\!\sim$ *satisfies the* **S**-*rule* (*i.e.*, $\alpha \wedge \beta \mid\!\sim \rho$ *implies* $\alpha \mid\!\sim \beta \to \rho$), *then* $\tilde{\rhd}$ *satisfies* **E-CM**.
(6) *If* $\mid\!\sim$ *satisfies* **WDR**, *then* $\tilde{\rhd}$ *satisfies* **LOR**.
(7) *If* $\mid\!\sim$ *is preferential and satisfies* **DR**, *then* $\tilde{\rhd}$ *satisfies* **E-DR**.
(8) *If* $\mid\!\sim$ *satisfies* **RM**, *then* $\tilde{\rhd}$ *satisfies* **E-R-Cut**.
(9) *If* $\mid\!\sim$ *is monotone, then* $\tilde{\rhd}$ *satisfies* **E-Cut**.

Remark.

(i) The hypothesis that $\mid\!\sim$ is adequate is only used to show $\mathbf{E\text{-}Con}_{\boldsymbol{\Sigma}}$ and **E-C-Cut**.
(ii) It is interesting to notice that we needed the **S**-rule, which is part of the preferential system, to get that $\tilde{\rhd}$ satisfies **E-CM** which is part of the cumulative system for explanatory relations.

Notice that Theorem 3.4 does not cover the case $\mid\!\sim$ cumulative. We will handle this case only for finite languages.

Proposition 3.5. *Suppose the language is finite. Let* $\mid\!\sim$ *be a cumulative relation such that* $\alpha \mid\!\sim \sigma$ *for all* α *and all* $\sigma \in \Sigma$. *Then there is an explanatory relation* $\rhd$ *satisfying*, **LLE**, **RLE**, **E-CM** *and* **E-C-Cut** *such that* $\mid\!\sim = \mid\!\sim_{ab}$.

3.1. Causal explanatory relations and reversed deduction

In the previous section we have shown that many consequence relations are of the form $\mid\!\sim_{ab}$. In this section we will address the dual question for explanatory relations. Namely, which explanatory relations are of the form $\tilde{\rhd}$? Let $\rhd$ be an explanatory relation and $\mid\!\sim_{ab}$

its associated consequence relation. Let $\tilde{\rhd}$ be the explanatory relation associated to $\mid\!\sim_{ab}$. Then the question is whether $\tilde{\rhd}$ is equal to $\rhd$. Consider the following condition on $\rhd$:

$$\alpha \rhd \gamma \text{ iff } C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \gamma). \tag{5}$$

Then our question can be equivalently stated as: Which explanatory relations satisfy (5)? First let us notice that in (5) the direction from left to right always holds. Condition (5) says that $\rhd$ can be recuperated from $\mid\!\sim_{ab}$ and thus explanatory reasoning based on $\rhd$ can be viewed as performing a sort of reversed deduction with respect to $\mid\!\sim_{ab}$. We will give more evidence about the last claim later in this section. The failure of (5) means that even if we know that an agent is reasoning abductively, we might not be sure which explanatory relation the agent is using. In other words, looking only at $\mid\!\sim_{ab}$ we cannot tell what are the agent's preferred explanations. We will isolate (5) in the following definition.

Definition 3.6. An explanatory relations is said to be *causal* if it satisfies (5).

In the following sections we will show some examples of explanatory relations which are far from being causal relations. Notice that $\mid\!\sim_{ab} = \vdash_\Sigma$ for any explanatory relation satisfying full reflexivity (i.e., $\alpha \rhd \alpha$ for every α consistent with Σ), thus such relations cannot be causal unless they are trivial.

So far we have not presented any semantic characterization of explanatory relations. It is not difficult to see that most causal explanatory relations can be easily characterized in terms of preferential models. Cumulative, preferential and rational relations are represented by cumulative, preferential and ranked models, respectively (see [6,10,12,18]). Those models can also be used to represent causal explanatory relations. In fact, from (5) it follows that one can check whether $\alpha \rhd \gamma$ holds by looking at the model that represents $\mid\!\sim_{ab}$. To give an example we state the theorem corresponding to E-rational causal relations.

Theorem 3.7. *Let $\rhd$ be a causal E-rational explanatory relation satisfying* **E-Con$_\Sigma$**. *Then there is a ranked preferential model $(mod(\Sigma), \preceq)$ such that for every γ consistent with Σ*

$$\alpha \rhd \gamma \textit{ iff } mod(\Sigma \cup \{\gamma\}) \subseteq min(\alpha).$$

Now we will address the question of when a relation is causal. The first observation is that any relation of the form $\tilde{\rhd}$ trivially satisfies **E-RW** and **RA**. We will need a bit more than these two postulates to get a characterization of causal relations.

Consider the following postulate:

C Let α and γ be formulas consistent with Σ. If for all δ such that $\delta \nvdash_\Sigma \bot$ and $\delta \vdash_\Sigma \gamma$ there is ρ such that $\alpha \rhd \rho$ and $\rho \vdash_\Sigma \delta$, then $\alpha \rhd \gamma$

This postulates says that if any consistent extension of γ can also be extended to a preferred explanation of α, then γ itself is a preferred explanation of α. Postulate **C** is a strong version of **E-RW** (in the presence of **RA**).

Proposition 3.8. *Let $\rhd$ be an explanatory relation. The following are equivalent*:

(i) $\rhd$ *is causal*;
(ii) $\rhd$ *satisfies* **RA** *and* **C**.

Next result shows that reversed deduction is a very particular form of causal explanatory reasoning. This result was essentially proved by Flach (he stated it differently, see Section 5).

Proposition 3.9. *Let $\rhd$ be an explanatory relation. The following are equivalent*:
(i) $\rhd$ *is causal and satisfies* **E-Cut** *and* **E-Con$_\Sigma$**;
(ii) $\alpha \rhd \gamma$ *iff* $\gamma \vdash_\Sigma \alpha$ *and* $\gamma \nvdash_\Sigma \bot$.

If the language is finite, causal explanatory relations are characterized by **RA** and **E-RW**. We will present a more general result that also applies to infinite languages. For that end we will require that every observation has at most finitely many preferred explanations. First, we introduce an auxiliary notion.

Definition 3.10. A set of formulas A is said to have an upper bound (in A with respect to Σ) if there are finitely many formulas $\alpha_1, \ldots, \alpha_n \in A$ such that for all $\alpha \in A$, $\alpha \vdash_\Sigma (\alpha_1 \vee \cdots \vee \alpha_n)$ (i.e., $\alpha_1 \vee \cdots \vee \alpha_n$ is an upper bound of A in the lattice of formulas modulo Σ).

Definition 3.11. An explanatory relation $\rhd$ is said to be *logically finite on the right* and denoted by **RLF**, if for every formula α the set $\{\gamma\colon\ \alpha \rhd \gamma\}$ has an upper bound.

Notice that if the language is finite then every explanatory relation obviously satisfies **RLF**.

Proposition 3.12. *Let $\rhd$ be an explanatory relation satisfying* **RA**, **E-RW** *and* **RLF**. *Then $\rhd$ is causal.*

We will show in Section 3.2 an example of a causal explanatory relations which does not satisfy **RLF**.

Corollary 3.13. *Suppose the language is finite and let $\rhd$ be explanatory relation. Then $\rhd$ is causal iff it satisfies* **E-RW** *and* **RA**.

What kind of relations are not causal? The examples that we will present in Section 4 use a notion of explanation based on belief revision which is a typical notion that does not satisfy **RA**.

3.2. More examples

It is easy to verified that the explanatory relations given in Section 2.1 are both causal. In fact, as the language is finite, both examples are of the form $\tilde{\rhd}$ for an adequate consequence relations $\mid\!\sim$. In Example 1 we have that $\mid\!\sim$ is preferential since we have used a partial order to define the preferential model and thus, by Theorem 3.4, $\tilde{\rhd}$ is E-preferential. In Example 2 the preference relation is a total pre-order and hence the consequence relations is rational and the associated explanatory relation is E-rational.

Example 3. This example is a minor modification of one given in [15]. Consider the following scenario: Lisa lives in a high-rise and parks her car in the 16-floor parking garage of her building. One morning, Lisa was looking for her car and did not find it where she thought she left it the night before. She considered the possibility that she was in the wrong floor and went to the next floor. There was also the possibility that the car was stolen and she must had called the police, but Lisa looked for the elevator and went to the next floor instead before taking the extreme decision of calling the police. We could model part of her background theory as follows: Let the language consist of the propositional variables $\{c, r, s, f, p\}$, where r stands for *right_floor*, c for *car*, s for *stolen_car*, f for *go_to_next_floor* and p for *call_police*. The background theory Σ will be the following:

$$\Sigma = \begin{cases} \neg r \to \neg c \\ \ \ s \to \neg c \\ \neg r \to \ \ f \\ \ \ s \to \ \ p \end{cases}$$

Lisa's preference are linearly pre-ordered. She prefers "worlds" where her car has not been stolen. In case the car is not found, she would think that she is not at the right place. So she has a three level preferential model:

$$\begin{aligned} L_0 &= \big\{\{r, c\}\big\}, \\ L_1 &= \big\{\{f\}, \{f, p\}\big\}, \\ L_2 &= \big\{\{r\}, \{r, p\}, \{r, f\}, \{r, c, f\}, \{r, c, p\}, \{r, s, p\}, \{r, f, p\}, \\ &\quad\ \{r, s, p, f\}, \{r, c, p, f\}, \{s, p, f\}\big\}. \end{aligned}$$

Notice that $mod(\Sigma) = L_0 \cup L_1 \cup L_2$. L_0 contains the initial states, in this case $\{r, c\}$. This is what Lisa expected before arriving to the parking place: the car will be there and she will not need to do anything else.

Let $\mid\!\sim$ be the rational consequence relation associated to this ranked model. That is to say

$$\alpha \mid\!\sim \beta \text{ iff } min(\alpha) \subseteq mod(\beta).$$

Let $\tilde{\rhd}$ be the explanatory relation associated to $\mid\!\sim$. Since the language is finite then $\mid\!\sim$ is adequate and by Theorem 3.4 we have that $\tilde{\rhd}$ is E-rational. Notice that $mod(\Sigma) = L_0 \cup L_1 \cup L_2$, hence $\tilde{\rhd}$ satisfies **E-Con$_{\boldsymbol{\Sigma}}$**. It is easy to check that the following holds:

$$\alpha \,\tilde{\rhd}\, \gamma \text{ iff } mod\big(\Sigma \cup \{\gamma\}\big) \subseteq min(\alpha).$$

We have that

$$mod\big(C(\neg c)\big) = \big\{\{f\}, \{f, p\}\big\}.$$

It is easy to check that

$$mod\big(\Sigma \cup \{\neg r\}\big) = \big\{\{f\}, \{f, p\}, \{s, p, f\}\big\}.$$

Thus $\neg c \not\tilde{\rhd} \neg r$, but it is clear that $\neg c \,\tilde{\rhd}\, (\neg r \wedge \neg s)$. So $\neg r$ is not enough to explain why the car was not found. Since $\neg r \wedge \neg s \mid\!\sim f$, then Lisa will go to the next floor. Notice also that $s \in Expla(\neg c)$, however $\neg c \not\tilde{\rhd} s$ because

$$mod\bigl(\Sigma \cup \{s\}\bigr) \not\subseteq mod\bigl(C(\neg c)\bigr)$$

(Lisa does not wish to think that the car was stolen as an explanation for not finding it). Observe also that $s \mid\!\sim \neg c$, so it is not sufficient that $\gamma \mid\!\sim \alpha$ in order that $\alpha \,\tilde{\rhd}\, \gamma$. Finally, to illustrate how $\tilde{\rhd}$ treats a disjunction, let us observe that $C(\neg c \vee s) = C(\neg c)$ and thus $(\neg c \vee s) \,\tilde{\rhd}\, (\neg r \wedge \neg s)$ but notice that $s \not\tilde{\rhd} (\neg r \wedge \neg s)$.

Example 4. This example is similar to Example 1 given in Section 2.1. Now we will allow messages of any length, but we will consider the situation of only one sender. Again the preference criterion is simple: messages starting with an even number of 0 are the most preferred ones. To make easier the presentation for each $n \geqslant 1$ let γ_n be the formula encoding the message of $2n + 1$ digits such that the first $2n$ digits are equal to 0 and the $(2n + 1)$th digit is equal to 1. Our language will be propositional on the countable set of variables $\{p_1, p_2, p_3, \ldots\}$ and Σ will be the empty set. [6] Let

$$L_0 = \bigcup_{n \geqslant 1} mod(\gamma_n)$$

and L_1 consists of all valuations not in L_0. We have then a two level ranked model. Let $\mid\!\sim$ be the rational consequence relation defined by this model and let $\tilde{\rhd}$ be explanatory relation associated with $\mid\!\sim$. It is not difficult to check that $\mid\!\sim$ is adequate and therefore by Theorem 3.4 we have that $\tilde{\rhd}$ is E-rational.

We will show that $\tilde{\rhd}$ is not logically finite on the right. In fact, suppose that the only portion of the message we were able to get only consists of ceros. Let us say

$$\alpha = \neg p_3 \wedge \neg p_5.$$

Then it is easy to check that $mod(\gamma_n) \subseteq min(\alpha)$ for all $n \geqslant 3$. Thus $\alpha \,\tilde{\rhd}\, \gamma_n$ for all $n \geqslant 3$ and therefore no preferred explanation of α is an upper bound for all preferred explanations of α. This shows that $\tilde{\rhd}$ is not logically finite on the right, but it is a causal explanatory relation by definition.

On the other hand, if the portion of the message contains at least one 1, then there is an upper bound for the set of preferred explanation for that message. For instance, let β (the incomplete message received) be $\neg p_2 \wedge p_5$. Then $\gamma_1 \wedge p_5$ and γ_2 are preferred explanations for β. In other words, the first five digits of the most likely messages sent are 00101, 00111 and 00001. In this case the upper bound is $(\gamma_1 \wedge p_5) \vee \gamma_2$.

Example 5. We will present examples of an adequate and nonadequate relation for an infinite language.

(i) Let $\{p_i\colon\ i \geqslant 1\}$ be the variables of the language and $\Sigma = \{p_1\}$. Consider the following two-level ranked preferential model: at the lowest level there will be only one model, M, defined by $M \models p_i$ for all $i \geqslant 1$ and at the second level we put all the other models of Σ (but not M). Let $\mid\!\sim$ be the relation associated with this ranked

[6] We could have put $\Sigma = \{\neg p_1\}$ to make this example closer to Example 1. But this is not important.

preferential model. Clearly $\mid\!\sim$ satisfies $\mathbf{Con}_\Sigma$. Let $\alpha = p_1$. It is clear that $C(\alpha) = Th(M)$, thus there is no γ (consistent with Σ) such that $C(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})$. Therefore (3) does not hold because its right hand side contains all formulas and its left hand side is equal to $Th(M)$.

(ii) Let Σ be the empty background theory and as in (i) we define a two-level ranked model: at the lowest level we put all models of p_1 and at the second level we put the other valuations of the language (i.e., those which do not satisfy p_1). Let $\mid\!\sim$ be the rational relation associated with this ranked preferential model. We claim that $\mid\!\sim$ is adequate. In fact, let α be any consistent formula. We consider two cases:
   (a) Suppose $\alpha \vdash \neg p_1$, then it is easy to check that $C(\alpha) = Cn(\alpha)$. From this it follows that (3) holds.
   (b) Suppose $\alpha \nvdash \neg p_1$, then it is easy to check that $C(\alpha) = Cn(\alpha \wedge p_1)$ and as before this implies that (3) holds.

4. Connection with belief revision

We will show in this section the connection of our approach with the theory of belief revision. In particular, we will see the peculiar place that causal explanatory relations occupy when they are viewed from the perspective of belief revision.

Belief revision is the process of changing the beliefs an agent has in order to incorporate incoming information (which might contradict the old one). The best known formalism for belief revision is the so called AGM postulates [1]. Let K be the belief set of an agent (which we assume to be a propositional theory) and suppose that the new incoming information is represented by a formula α. The revision of K with α is denoted by $K * \alpha$. It is natural to assume that $K * \alpha$ is also a belief set (i.e., closed under logical consequences) and obviously that $\alpha \in K * \alpha$. The AGM postulates impose other nontrivial conditions on $*$ in order to make minimal the changes it performs in K. For instance, if α is consistent with K then $K * \alpha = Cn(K \cup \{\alpha\})$. Gärdenfors and Makinson [9] have shown a tight connection of belief revision with the theory of nonmonotonic consequence relations. Given an AGM revision operator $*$ they define a consequence relation by letting $\alpha \mid\!\sim_K \beta$ if $\beta \in K * \alpha$. In words, its says that the agent is willing to conclude β from α in the case that β belongs to the revised belief set obtained after α is incorporated into K (using the revision operator $*$). In [8] it is shown that $\mid\!\sim_K$ is a rational consequence relation in the sense of Kraus et al. [10]. On the other hand, they also have shown that every rational consequence relation $\mid\!\sim$ can be represented as a consequence relation of the form $\mid\!\sim_K$. In fact, let $\mid\!\sim$ be a rational consequence relation and let

$$K = \{\alpha\colon\ \top \mid\!\sim \alpha\}.$$

Define $*$ by $K * \alpha = C(\alpha)$. Then $*$ is a revision operator for K such that $\mid\!\sim$ is equal to $\mid\!\sim_K$.[7]

[7] Formally $*$ cannot be considered a revision operator because we have given only a description of how to revise a single knowledge base, namely $C(\top)$, and $*$ must be applicable to any knowledge base. Also $*$ might not satisfy one of the defining condition of an AGM operator. Namely, $*$ might not preserve consistency: It can happen that α is consistent but $K * \alpha$ is inconsistent. To avoid this problem one has to restrict to rational consequence relations that preserve consistency: $\alpha \nvdash \bot$ iff $\alpha \mid\!\not\sim \bot$.

The connection between abduction and belief revision was already observed by Gärdenfors [7]. Boutilier and Becher [3] proposed a model of abduction based on the revision of the epistemic state of an agent. Aliseda [2] consider modeling belief revision with abduction (see also [14]). The main idea in all these papers is the same. We will follow the terminology of [3]. They consider various forms of explaining α relative to K and to an arbitrary (but fix) AGM revision operator. These type of explanations were called *epistemic explanations*. Epistemic explanations capture the intuition that *if the explanation were believed, so too would be the observation*. More precisely, they introduced the following.

Definition 4.1.[8] Let $*$ be an AGM revision operator and K be a consistent set of formulas. An *epistemic explanation* for α relative to K and $*$ is any consistent formula γ such that $\alpha \in K * \gamma$.

It is not difficult to see that the notion of epistemic explanations does not satisfy the postulate **RA**. Because if γ is an epistemic explanation of α, then $\gamma \wedge \delta$ is not in general an epistemic explanation of α. The reason is that $K * (\gamma \wedge \delta)$ is in general very different from $K * \gamma$. These notions of epistemic explanations "cannot be given a truly causal interpretation because they are simple beliefs that induce belief in the fact to be explained" [3]. The lack of a causal relationship between an observation and its epistemic explanations is precisely where our notion of causal explanation differs from theirs. There is also another very important difference. The relation "γ is an epistemic explanation of α" is not an explanatory relation in our sense. This is simply because an epistemic explanation might not have any deductive relationship with the explanandum. However, as revision operator preserves consistency, it is easy to see that an epistemic explanation has to be at least consistent with the explanandum.[9] We will make a little detour in order to introduce a new concept that covers the notion of epistemic explanations.

Definition 4.2. A binary relation $|\!<$ is called a *weak explanatory relation* if for all α and γ

$$\alpha |\!< \gamma \Rightarrow \gamma \wedge \alpha \not\vdash_{\Sigma} \bot .$$

Remark. Observe that for a weak explanatory relation its associate consequence relation $\mid\!\sim_{ab}$ is not necessarily reflexive. Thus $\mid\!\sim_{ab}$ might lose one of its more basic features and therefore it is not clear the role that $\mid\!\sim_{ab}$ could play for studying weak explanatory relations. All postulates we have introduced in Section 2 also apply to weak explanatory relations. Some of the results proved for explanatory relations are valid for weak explanatory relations. For instance Proposition 2.17 is valid. The proof of Proposition 2.10 works for weak explanatory relation, so **E-DR** implies **LOR** in this case too. It is easy to check that any weak explanatory relation satisfying **RA** is necessary an explanatory relation.

[8] This definition corresponds to what Boutilier and Becher called *predictive explanations*. This notion is the closer to our approach. We will not analyze other alternatives.

[9] We are assuming here that Σ is the empty set. This is not a crucial assumption. Our claims can easily extended to cover the case where Σ is not empty.

Let's go back to the main theme of this section. Recall the rational consequence relation $\mid\!\sim_K$ associated to an AGM revision operator. The notion of epistemic explanation can then be restated as follows:

γ is an epistemic explanation for α iff $\gamma \mid\!\sim_K \alpha$.

From this it is obvious what are the logical properties satisfied by epistemic explanations. However, it is convenient to see which of our postulates for explanatory reasoning are satisfied by epistemic explanations.

Proposition 4.3. *Assume that Σ is the empty set. Let $*$ be an AGM revision operator and K be a consistent set of formulas. Let $\mid\!<$ be defined by $\alpha \mid\!< \gamma$ if γ is consistent and $\alpha \in K * \gamma$. Then $\mid\!<$ is a weak explanatory relation that satisfies* **LLE**, **RLE**, **E-CM**, **E-RW**, **ROR**, **LOR**, **E-Cut** *and full reflexivity* (*i.e.*, $\alpha \mid\!< \alpha$ *for all consistent* α).

Epistemic explanation are far from being causal in our sense, since **RA** does not hold. Also let us remark that since transitivity of $\mid\!\sim$ implies monotonicity, then the notion of epistemic explanation is not transitive. [10]

The notion of epistemic explanation is too permissive. We can restrict it by asking a bit more from the explanations. Namely, we will say that γ is a *strong epistemic explanation* of α if

$$K * \alpha \subseteq K * \gamma. \tag{6}$$

In other words, after revising K with the explanation we obtain all beliefs corresponding to the revision of K with the observation. If we state this new definition in terms of $\mid\!\sim_K$ we get the following condition: $C_K(\alpha) \subseteq C_K(\gamma)$. Where, as usual, $C_K(\alpha) = \{\beta\colon \alpha \mid\!\sim_K \beta\}$. It is convenient to see this condition as defining a notion of an explanation with respect to an arbitrary consequence relation $\mid\!\sim$. More precisely, consider the following condition for any γ such that $\gamma \not\mid\!\sim \bot$

$$C(\alpha) \subseteq C(\gamma). \tag{7}$$

This condition was suggested by Flach (Lehmman [11] has some preliminaries results about it [11]). In our setting it is quite natural to require that $\mid\!\sim$ satisfies $\mathbf{Con}_\Sigma$. The next theorem shows which postulates are satisfied by this weak explanatory relation.

Proposition 4.4. *Let $\mid\!\sim$ be a preferential consequence relation satisfying* $\mathbf{Con}_\Sigma$. *Define $\alpha \mid\!< \gamma$ if* (7) *holds for γ consistent with Σ. Then $\mid\!<$ is a weak explanatory relation and moreover*:

(i) $\mid\!<$ *is transitive, full reflexive for Σ-consistent formulas and satisfies* **LLE**, **RLE**, **E-CM**, **E-RW** *and* **E-C-Cut**.

---

[10] We should mention that the original definition of predictive explanation given by Boutilier and Becher requieres an additional condition. When the observation α is entailed by K then they ask also that $\neg\gamma \in K * \neg\alpha$ which captures the intuition that if the observation had been absent, so too would be the explanation. With this extra restriction we have that **E-C-Cut** holds but we do not have neither **E-Cut** nor **E-R-Cut**.

[11] We thank him for letting us have a copy of his manuscript.

(ii) *If in addition* $\mid\!\sim$ *satisfies* **DR**, *then* $\mid\!<$ *satisfies* **LOR**.
(iii) *If in addition* $\mid\!\sim$ *satisfies* **RM**, *then* $\mid\!<$ *satisfies* **E-DR**, **ROR** *and* **E-R-Cut**.

Note that the relation $\mid\!<$ (given in Proposition 4.4) satisfies **E-Cut** iff $\mid\!\sim$ is monotonic. This relation is also far form being causal, since **RA** does not hold.

Since $K * \gamma$ is supposed to be closed under logical consequences and in our setting $\Sigma \subseteq K * \gamma$, then we have that $Cn(\Sigma \cup \{\gamma\}) \subseteq K * \gamma$. This suggests another way of strengthening (6). Consider the following notion of explanation

$$K * \alpha \subseteq Cn(\Sigma \cup \{\gamma\}). \tag{8}$$

This is exactly the defining condition of a causal explanatory relation. Let us see this in detail.

Let $\rhd$ be a causal explanatory relation. This means that the following holds

$$\alpha \rhd \gamma \text{ iff } C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \gamma). \tag{9}$$

Suppose also that $\rhd$ is E-rational and satisfies $\mathbf{E\text{-}Con}_{\Sigma}$. Then by Theorem 2.12 we know that $\mid\!\sim_{ab}$ is a rational consequence relation satisfying $\mathbf{Con}_{\Sigma}$. As before, let $*$ be the revision operator associated with $\mid\!\sim_{ab}$.[12] By definition $C_{ab}(\alpha)$ is equal to $K * \alpha$ and thus from (9) we have the following

$$\alpha \rhd \gamma \text{ iff } K * \alpha \subseteq Cn(\Sigma \cup \gamma)$$

which is exactly (8).

The initial knowledge base K is the collection $\{\alpha\colon \top \mid\!\sim_{ab} \alpha\}$. That is to say

$$K = \bigcap \bigl\{Cn(\Sigma \cup \{\gamma\})\colon \top \rhd \gamma\bigr\}.$$

K represents the agent's belief before any observation is made. It is clear that $\Sigma \subseteq K$ and moreover, by $\mathbf{E\text{-}Con}_{\Sigma}$, we have also that $\Sigma \subseteq K * \alpha$ for all α. Thus, after an observation is made, the belief set K is revised without modifying Σ. It is not hard to check (using **RA**) that an observation α is consistent with K iff there is γ such that $\top \rhd \gamma$ and $\alpha \rhd \gamma$.

To give a precise interpretation of (8) we must consider the following condition

$$Cn(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\}). \tag{10}$$

This corresponds to the notion of explanation given by Flach's postulates [5] and as we have proven in Proposition 3.9 it also corresponds to causal explanatory relations satisfying $\mathbf{E\text{-}Con}_{\Sigma}$ and **E-Cut**. It is clear that (10) can be viewed as performing an expansion of the knowledge base [1,7] instead of a revision.

Notice that (8) is stronger than (6). Thus any preferred explanation is a strong epistemic explanation. However, rather than saying that γ normally implies α (as Boutilier and Becher did) we say that γ implies everything that is normally implied by α. Condition (8) keeps some of the "epistemic" flavor of the belief revision approach and at the same time retains a strong causal relationship between an observation and its preferred explanations. Causal explanatory relations treats differently observations and explanation.

[12] As we said before, $*$ is not formally an AGM revision operator. However, it still captures the key idea of belief revision, that is to say, to minimize the changes of K.

An observation has associated some beliefs (the other "symptoms" that we believe usually are also present) so we could say loosely that observations are treated as beliefs. However, explanations are not treated as beliefs. This epistemological distinction seems to capture the following idea. *We might be wrong about which is the "real world" (i.e., the preference relation might be incorrect), but we would like to be right about the causality relation used to explain the features of whichever world we happen to prefer*.

Example. To illustrate the differences between epistemic, strong epistemic and causal explanations let's go back to Lisa's example in Section 3.2. In this example K is the theory of $\{r, c\}$ which correspond to what Lisa expected before arriving to the parking place. An AGM revision operator is defined in the usual way: $K * \alpha$ corresponds to the theory of the minimal models of α (with respect to the total pre-order of $mod(\Sigma)$ given in Section 3.2).

It is easy to verify that f is a strong epistemic explanation of $\neg c$ (but notice that $f \wedge r$ is not). However, for us f is not even an explanation of $\neg c$ since $\Sigma \cup \{f\} \nvdash \neg c$. Another instance, $\neg r$ is a strong epistemic explanation of $\neg c$, it entails $\neg c$ but it is not a preferred explanation in our sense. On the other hand, $\neg r \wedge \neg s$ is both a preferred explanation and a strong epistemic explanation of $\neg c$. Finally, $r \wedge p \wedge \neg c$ is an epistemic explanation of $\neg c$ but it is not a strong epistemic explanation of $\neg c$.

5. Related works

We will comment in this section about the connection of our results and the work of Flach [5], Cialdea–Pirri [4], Aliseda [2], Lobo–Uzcátegui [15] and Zadrozny [20].

P. Flach

His work is the closest to ours. He presented some postulates for explanatory and inductive reasoning. Some of our postulates are similar to his. He studied the relations "γ is a possible inductive hypothesis given evidence α" and "γ is a possible explanation of α" which he denotes by $\alpha |\!< \gamma$. He did not assume that $|\!<$ is an explanatory relation, however one of his postulates implies that $|\!<$ has to be weak explanatory. Flach uses a satisfaction relation $\models$ instead of $\vdash_\Sigma$ and thus the background theory is not mentioned explicitly. Below we will compare his postulates with ours.

I1. If $\alpha |\!< \gamma$ and $\models \alpha \wedge \gamma \to \beta$, then $(\alpha \wedge \beta) |\!< \gamma$. When $|\!<$ is assumed to be an explanatory relation then it is not difficult to see that **I1** is, in our context, equivalent to **E-CM**.

I2. If $\alpha |\!< \gamma$ and $\models \alpha \wedge \gamma \to \beta$, then $(\alpha \wedge \neg\beta) |\!\not< \gamma$. This says that $|\!<$ is a weak explanatory relation.

I3. If $\alpha |\!< \gamma$ and $\models \alpha \wedge \gamma \to \beta$, then $\alpha |\!< \gamma \wedge \beta$. When $|\!<$ is assumed to be an explanatory relation **I3** follows from **RLE**.

He considered two versions of Reflexivity:

I4. If $\alpha |\!< \gamma$, then $\alpha |\!< \alpha$. This postulate will not be valid in general in our case, because α might not be a *preferred* explanation of itself.

I5. If $\alpha|\!<\gamma$, then $\gamma|\!<\gamma$. We already have mentioned that **I5** holds for explanatory relations satisfying **E-CM** and **E-C-Cut**.

The other two postulates for induction **I6** and **I7** correspond to **RLE** and **LLE**, respectively. Flach studied other postulates more specific of explanatory reasoning:

E1. If $\alpha|\!<\delta$, $\gamma|\!<\gamma$ and $\models \gamma \to \delta$, then $\alpha|\!<\gamma$. This postulate is essentially **RA**.

E2. If $\gamma|\!<\gamma$ and $\neg\alpha|\!\not<\gamma$, then $\alpha|\!<\alpha$. This postulate does not necessarily hold in our case. In our context, this rule is quite strange because it says that when a formula α is not a good explanation for itself then any good explanation is a good explanation for the negation of α. This rule will be valid in the monotonic case.

E3. If $\alpha|\!<(\beta \wedge \gamma)$, then $(\beta \to \alpha)|\!<\gamma$. This rule seems to be valid only in the monotonic case.

E4. If $\alpha|\!<\gamma$ and $\beta|\!<\gamma$, then $(\alpha \wedge \beta)|\!<\gamma$. This postulate is a consequence of **E-CM**.

E5. If $\alpha|\!<\gamma$ and $\models \alpha \to \beta$, then $\beta|\!<\gamma$. This postulate implies **E-Cut** and in fact, it is equivalent to **E-Cut** under the presence of **E-CM**.

He then presented five postulates for "confirmatory induction" which does not seem to be applicable for explanatory reasoning, except for his postulate C4 which corresponds to our **LOR**. For Flach "intuition constitutes the primary source of justification for his rationality postulates". Our results confirm that his intuition also has a formal justification. The more important difference with our approach is that he did not consider weaker cut rules than **E-Cut** thus his postulates force $\mid\!\sim_{ab}$ to be monotonic. Moreover, his main representation theorem for explanatory relations says that explanatory reasoning is restricted to reversed deduction. More formally, he showed the following

Theorem. *A binary relation* $|\!<$ *satisfies* **I1–7** *and* **E1–5** *iff the following holds*:

$$\alpha|\!<\gamma \textit{ iff } \models \gamma \to \alpha.$$

We will sketch a proof of this result based on Proposition 3.9. Flach's formalism does not explicitly include a postulate similar to **E-Con$_\Sigma$**. However, it follows from his postulates that for a carefully chosen Σ our **E-Con$_\Sigma$** holds (showing this fact is in part what makes Flach's proof long). We will assume that **E-Con$_\Sigma$** holds and use $\vdash_\Sigma$ instead of $\models$. Notice, that by **I4**, **I5** and **E-Con$_\Sigma$** we have that a formula is admissible iff it is consistent with Σ. First, one has to show that $|\!<$ is an explanatory relation. In fact, by **E3** and **E-Cut** one gets that $\alpha|\!<\gamma$ iff $(\gamma \to \alpha)|\!<\top$ and from this it is not difficult to show using **E-Con$_\Sigma$** that $|\!<$ is an explanatory relation. In order to use Proposition 3.9 it suffices to verify postulate **C**. Suppose that γ and α are consistent with Σ and the hypothesis in postulate **C** hold. First, using the hypothesis in **C** we have that $\gamma \wedge \alpha$ is consistent with Σ. Then, assuming towards a contradiction that $\alpha|\!\not<\gamma$, we have that $(\gamma \to \alpha)|\!\not<\top$. From **E5** and **I4** it follows that $\top|\!<\top$. Now apply **E2** and get $\gamma \wedge \neg\alpha|\!<\top$, thus $\gamma \wedge \neg\alpha$ is consistent with Σ. Finally, using again the hypothesis in **C** one gets ρ such that $\rho \vdash_\Sigma \gamma \wedge \neg\alpha$ and $\alpha|\!<\rho$, from which one gets a contradiction.

Cialdea–Pirri

They defined a relation $\Sigma \vdash \gamma \rightsquigarrow \alpha$ to capture the notion that "in the theory Σ, γ is a good reason for α". The definition of $\rightsquigarrow$ is based on a preference relation over formulas as follows. Let $\prec$ be an irreflexive relation on formulas. The explanatory relation $\rhd$ associated with $\prec$ is defined by:

$$\alpha \rhd \gamma \Leftrightarrow \gamma \in min(Expla(\alpha), \prec). \tag{11}$$

In other words, $\alpha \rhd \gamma$ iff $\nvdash_\Sigma \neg\gamma$, $\gamma \vdash_\Sigma \alpha$ and $\delta \nvdash_\Sigma \alpha$ for all δ such that $\delta \prec \gamma$. It is easy to check that such explanatory relations always satisfy **E-Reflexivity** and **E-CM**.

They presented some basic postulates and some conditions where they hold. Our postulate **E-CM** is stronger than their **And-Right**. Their Left Logical Equivalence is our **RLE**. Our Cut rules (**E-C-Cut**, **E-R-Cut** and **E-Cut**) have nothing to do with their E-Cut. Here there is an important difference between our approach and theirs. As we said in the introduction, we consider the background theory Σ fixed, but they considered postulates concerning properties of abduction when the background theory changes. For instance, their E-Cut rule says

If $\Sigma \vdash \alpha$ and $\Sigma \cup \{\alpha\} \vdash \gamma \rightsquigarrow \beta$, then $\Sigma \vdash \gamma \rightsquigarrow \beta$.

and their E-Monotonicity rule says

If $\Sigma \vdash \alpha$ and $\Sigma \vdash \gamma \rightsquigarrow \beta$, then $\Sigma \cup \{\alpha\} \vdash \gamma \rightsquigarrow \beta$.

These last two postulates are very weak, since they are valid for every explanatory relation $\rhd$ defined as in (11) regardless of the preference relation $\prec$ used. They did not study the problem of whether their postulates will guarantee that $\rightsquigarrow$ is given by a preference relation (this will be addressed in [19]).

Atocha Aliseda

Her Ph.D. Thesis is a comprehensive presentation of abduction from several points of view. It is a very good source for the vast literature on abduction. We will make some comments only about the part of her work which is close related to our paper. Similar to Cialcea and Pirri's approach, Aliseda regards abduction as a relation with three parameters: a background theory, an observation and an explanation. Her notation is $\Sigma \mid \gamma \Rightarrow \alpha$ to express that γ is an explanation for α with respect to Σ. She presented sets of rules for various versions of abduction: Plain, Consistent, Explanatory, Minimal and Preferential abduction. Some of her postulates are not valid in our context, for instance her Weak Explanatory Reflexivity says

If $\Sigma \mid \gamma \Rightarrow \alpha$, then $\Sigma \mid \alpha \Rightarrow \alpha$

which is Flach's **I4** and, as we already said, it is not valid in our context because in most cases an observation is not a preferred explanation of itself. She also consider cut and monotonicity rules similar to those used by Cialcea and Pirri. However, no cut rule for observations (as ours) was studied, except the rule of transitivity (which follows from **RA**). Among all versions of abduction she considered, Preferential abduction is the closest to our approach. It naturally requires that γ has to be minimal with respect to a preference relation among formula. The crucial rule for axiomatizing Preferential abduction is the following:

If $\Sigma \mid \gamma \Rightarrow \alpha$ and $\gamma \wedge \delta < \gamma$, then $\Sigma \mid \gamma \wedge \delta \Rightarrow \alpha$

where $<$ is a preference order among formulas. Aliseda does not view this rule as structural rule since it requieres a preference relation that she thought cannot be expressed in terms of the inference relation itself. But we have shown in [19] that preference criteria can be coded by the structural rules without explicitly mention them.

Lobo–Uzcátegui

In logic-based abduction usually together with the background theory Σ there is also a distinguished set of atoms *Ab* called *abducibles*. Formulas using only atoms from *Ab* are also called abducible. The pair (Σ, Ab) is referred to as the *Abductive framework*. Let $\mid\!\sim$ be a consequence relation satisfying $\mathbf{Con}_\Sigma$. An *Ab-explanation* of α is any abducible formula γ consistent with Σ such that $\gamma \mid\!\sim \alpha$. Thus this notion of explanation is similar to the notion of epistemic explanation. Assuming the language is finite, the *cautious explanation* of α, denoted by $F_c(\alpha)$, is defined as the disjunction of all *Ab*-explanations of α. Define $\mid\!\sim_a$ by letting $\alpha \mid\!\sim_a \beta$ if $F_c(\alpha) \mid\!\sim \beta$. This type of consequence relations $\mid\!\sim_a$ (especially when $\mid\!\sim$ is rational) were studied in [15]. Notice that if $\gamma \mid\!\sim \beta$ for all *Ab*-explanation γ of α, then $\alpha \mid\!\sim_a \beta$. However, the converse is not true because in the definition of $\mid\!\sim_a$ there is an implicit selection of some Ab-explanation of α as the preferred ones. The relation $\mid\!\sim_{ab}$ introduced in this paper was motivated by $\mid\!\sim_a$. The role of abducibles formulas in [15] is quite closed to our admissible formulas.

W. Zadrozny

He approached abduction from a quite abstract point of view based on the concept of invariant of reasoning. Abduction is viewed as an inference process that preserves sets of explanations. It is not clear the relation with our results, but it seems an interesting topic of research. He has some rules similar to ours but his presentation is quite complex. His explanation systems are formulated using higher-order logic as a metalanguage.

6. Conclusions

We have analyzed two aspects of explanatory reasoning: Its logical properties and its relation with reversed deduction. The logical properties have been isolated in a fairly complete list of postulates. Some of our postulates are similar to some of those introduced by previous approaches (Flach, Cialdea–Pirri and Aliseda). The key idea was to use $\mid\!\sim_{ab}$ as an heuristic device for isolating the logical properties of an explanatory relation $\rhd$. It is important to point out the special role that explanatory cut rules play in our presentation. We have not seen these rules in other formalism.

When we started this research we were focused on getting $\mid\!\sim_{ab}$ to have good properties in the KLM sense. Moreover, we thought that an explanatory relation $\rhd$ and its associated consequence relation $\mid\!\sim_{ab}$ were somewhat interchangeable. But this turns out to be true only for those explanatory relations that we have called *causal*. For a noncausal explanatory relation there is a lost of information when going from $\rhd$ to $\mid\!\sim_{ab}$. Because in this case, even if we know that an agent is reasoning abductively, we might not be sure which explanatory relation the agent is using. In other words, looking only at $\mid\!\sim_{ab}$ we cannot tell what are the agent's preferred explanations.

We have shown that causal explanatory reasoning is nonmonotonic reasoning in reverse. This answers one of our initial questions. However, it is important to remark a difference between the postulates for explanatory relations and nonmonotonic consequence relations. The basic postulates (in the KLM style) for nonmonotonic consequence relations can be stated as inference rules in a propositional language, but for explanatory relation some of the basic postulates (like **E-C-Cut** and **C**) are expressed as first-order properties of $\rhd$.

Causal explanatory relations have also a interpretation in terms of belief revision. The key feature that distinguishes causal explanations from other notions of explanations is the fact that causal explanatory relations treat observations and explanations in a different way. An observation has associated some beliefs (the other "symptoms" that we believe usually are also present) so we could say loosely that observations are treated as beliefs. However, explanations are not treated as beliefs and the deductive relationship between an observation and its preferred explanations is retained in a very strong form. The underlying idea of causal explanatory relations is the following. After observing α, we first collect the concomitant facts that are normally present (i.e., we compute $C_{ab}(\alpha)$) and then we select the preferred explanations of α as those formulas that entails α and its usual consequences $C_{ab}(\alpha)$. In other words, rather than saying that γ normally implies α we say that γ implies everything normally implied by α.

Finally, we will mention two possible lines of research related to our results. The first one is to study more carefully the hierarchy we have presented for classifying the logical properties of abduction. Specially relevant is to determine up to which extend this hierarchy classifies (noncausal) weak explanatory relations. The second one is related to the role of the background theory. Usually it is said there are three kinds of reasoning processes: deductive, abductive and inductive. We have shown that abduction is very tightly related to a "nonmonotonic-deduction". On the other hand, inductive reasoning (when it is understood as the process of inferring general rules out of specific observations) did not play any role in our setting. This is probably due to the fact that we have fixed the background theory. There are many situations where Σ is the natural outcome of an inductive reasoning process. As we said in the introduction, Cialdea–Pirri and Aliseda presented a view of abduction as a relation with three parameters: an observation, an explanation and a background theory. We think that an extension of our results, to the more general case where the background theory is allowed to change, will provide some hints for a better understanding of inductive reasoning.

Acknowledgements

We would like to thank Jorge Lobo, M. Freund, D. Lehmann and the anonymous referees for their comments and criticism which have improved the quality of the paper.

Partial support for Carlos Uzcátegui was provided by a CDCHT-ULA (Venezuela) grant. This work was initiated in 1996 while he was visiting the Laboratoire d'Informatique Fondamentale de Lille (LIFL), France. He would like to thank LIFL for the financial assistance and facilities they provided.

Appendix A. Rationality postulates for consequence relations

To make easier the reading of the paper we will include a list of all rationality postulates for consequence relations used in the paper.

REF (Reflexivity)	$\alpha \mid\!\sim \alpha$
LLE (Left Logical equivalence)	$\alpha \mid\!\sim \beta \;\&\; \vdash \alpha \leftrightarrow \gamma \;\Rightarrow\; \gamma \mid\!\sim \beta$
RW (Right Weakening)	$\alpha \mid\!\sim \beta \;\&\; \vdash \beta \rightarrow \gamma \;\Rightarrow\; \alpha \mid\!\sim \gamma$
CUT	$\alpha \wedge \beta \mid\!\sim \gamma \;\&\; \alpha \mid\!\sim \beta \;\Rightarrow\; \alpha \mid\!\sim \gamma$
CM (Cautious Monotony)	$\alpha \mid\!\sim \beta \;\&\; \alpha \mid\!\sim \gamma \;\Rightarrow\; \alpha \wedge \gamma \mid\!\sim \beta$
OR	$\alpha \mid\!\sim \gamma \;\&\; \beta \mid\!\sim \gamma \;\Rightarrow\; \alpha \vee \beta \mid\!\sim \gamma$
S	$\alpha \wedge \beta \mid\!\sim \gamma \;\Rightarrow\; \alpha \mid\!\sim \beta \rightarrow \gamma$
DR (Disjunctive Rationality)	$\alpha \vee \beta \mid\!\sim \rho \;\Rightarrow\; \alpha \mid\!\sim \rho$ or $\beta \mid\!\sim \rho$
RM (Rational Monotony)	$\alpha \mid\!\sim \rho \;\&\; \alpha \not\mid\!\sim \neg\beta \;\Rightarrow\; \alpha \wedge \beta \mid\!\sim \rho$
Mono (Monotony)	$\alpha \mid\!\sim \gamma \;\Rightarrow\; \alpha \wedge \beta \mid\!\sim \gamma$

An inference relation $\mid\!\sim$ is said to be *cumulative* if it satisfies the rules **REF**, **LLE**, **RW**, **CUT** and **CM**. A consequence relation is called *preferential* if it satisfies, in addition to cumulative rules, the rule **OR** and it is called *rational* if it is preferential and satisfies **RM**. $\mid\!\sim$ is *monotone* if it satisfies **Mono**. A consequence relation satisfies **WDR** if $C(\alpha \vee \beta) \subseteq Cn(C(\alpha) \cup C(\beta))$, for every formulas α and β. We used also **Con**$_{\Sigma}$ (Σ-consistency preservation) which is a variant of a postulate introduced in [9]: for all α, $\alpha \mid\!\sim \bot$ iff $\vdash_{\Sigma} \neg\alpha$ and if $\sigma \in \Sigma$, then $\alpha \mid\!\sim \sigma$.

Appendix B. Rationality postulates for explanatory relations

We list below all postulates for explanatory relations that we have introduced in this paper.

LLE	$(\vdash_{\Sigma} \alpha \leftrightarrow \alpha') \;\&\; \alpha \rhd \gamma \;\Rightarrow\; \alpha \rhd \gamma'$
RLE	$(\vdash_{\Sigma} \gamma \leftrightarrow \gamma') \;\&\; \alpha \rhd \gamma \;\Rightarrow\; \alpha \rhd \gamma'$
E-CM	$\alpha \rhd \gamma \;\&\; \gamma \vdash_{\Sigma} \beta \;\Rightarrow\; (\alpha \wedge \beta) \rhd \gamma$
E-Cut	$(\alpha \wedge \beta) \rhd \gamma \;\Rightarrow\; \beta \rhd \gamma$
E-C-Cut	$(\alpha \wedge \beta) \rhd \gamma \;\&\; \forall\delta\, (\alpha \rhd \delta \Rightarrow \delta \vdash_{\Sigma} \beta) \;\Rightarrow\; \alpha \rhd \gamma$
E-R-Cut	$(\alpha \wedge \beta) \rhd \gamma \;\&\; \exists\delta\, [\alpha \rhd \delta \;\&\; \delta \vdash_{\Sigma} \beta] \;\Rightarrow\; \alpha \rhd \gamma$

E-Reflexivity $\alpha \rhd \gamma \Rightarrow \gamma \rhd \gamma$

E-RW $\alpha \rhd \gamma \;\&\; \alpha \rhd \delta \Rightarrow \alpha \rhd (\gamma \vee \delta)$

ROR $\alpha \rhd \gamma \vee \rho \Rightarrow \alpha \rhd \gamma$ or $\alpha \rhd \rho$

E-Disj $\gamma \nvdash_\Sigma \bot \;\&\; \rho \nvdash_\Sigma \bot \;\&\; \alpha \rhd (\gamma \vee \rho) \Rightarrow \alpha \rhd \gamma \;\&\; \alpha \rhd \rho$

RA $\alpha \rhd \gamma \;\&\; \gamma' \vdash_\Sigma \gamma \;\&\; \gamma' \nvdash_\Sigma \bot \Rightarrow \alpha \rhd \gamma'$

LOR $\alpha \rhd \gamma \;\&\; \beta \rhd \gamma \Rightarrow (\alpha \vee \beta) \rhd \gamma$

E-DR $\alpha \rhd \gamma \;\&\; \beta \rhd \delta \Rightarrow (\alpha \vee \beta) \rhd \gamma$ or $(\alpha \vee \beta) \rhd \delta$

E-Con$_\Sigma$ $\nvdash_\Sigma \neg\alpha \Leftrightarrow \exists \gamma\; \alpha \rhd \gamma$

C Let α and γ be formulas consistent with Σ. If for all δ such that $\delta \nvdash_\Sigma \bot$ and $\delta \vdash_\Sigma \gamma$ there is ρ such that $\alpha \rhd \rho$ and $\rho \vdash_\Sigma \delta$, then $\alpha \rhd \gamma$.

Appendix C. Summary of the main results in Sections 2 and 3

See Tables C.1 and C.2.

Table C.1
From explanatory relations to consequence relations

$\rhd$		$\mid\!\sim_{ab}$
	$\Rightarrow$	Adequate + **REF** + **RW**
E-Con$_\Sigma$	$\Rightarrow$	**Con$_\Sigma$**
LLE + **E-CM** + **E-C-Cut**	$\Rightarrow$	Cumulative
LLE + **E-CM** + **E-C-Cut** + **RA**	$\Rightarrow$	Preferential
LLE + **E-CM** + **E-C-Cut** + **RA** + **LOR** + finite language	$\Rightarrow$	Preferential + **WDR**
LLE + **E-CM** + **E-C-Cut** + **RA** + **E-DR**	$\Rightarrow$	Preferential + **DR**
LLE + **E-CM** + **E-C-Cut** + **RA** + **E-R-Cut**	$\Rightarrow$	Rational
E-Cut	$\Rightarrow$	Monotonic

Table C.2
From consequence relations to explanatory relations

$\tilde{\rhd}$		$\mid\!\sim$ adequate
RA + **RLE** + **E-RW**	$\Leftarrow$	
E-Con$_\Sigma$	$\Leftarrow$	**Con$_\Sigma$**
LLE + **E-CM** + **E-C-Cut** + **RA** + **RLE** + **E-RW**	$\Leftarrow$	Preferential
LLE + **E-CM** + **E-C-Cut** + **RA** + **LOR** + **RLE** + **E-RW**	$\Leftarrow$	Preferential + **WDR**
LLE + **E-CM** + **E-C-Cut** + **RA** + **E-DR** + **RLE** + **E-RW**	$\Leftarrow$	Preferential + **DR**
LLE + **E-CM** + **E-C-Cut** + **RA** + **E-R-Cut** + **RLE** + **E-RW**	$\Leftarrow$	Rational
E-Cut + **RA** + **RLE** + **E-RW**	$\Leftarrow$	Monotonic

Appendix D. Proofs

Proof of Proposition 2.3. Suppose $\alpha \mathrel{|\!\sim}_{ab} \rho$, i.e., for all γ if $\alpha \rhd \gamma$ then $\gamma \vdash_\Sigma \rho$. Let δ be any formula such that $(\alpha \wedge \beta) \rhd \delta$. By **E-Cut** we have $\alpha \rhd \delta$ so $\delta \vdash_\Sigma \rho$. Thus $(\alpha \wedge \beta) \mathrel{|\!\sim}_{ab} \rho$. $\square$

Proof of Theorem 2.5. Suppose $\rhd$ is a relation as in the hypothesis. We will show that $\mathrel{|\!\sim}_{ab}$ is cumulative. From **LLE** for $\rhd$ we easily get that $\mathrel{|\!\sim}_{ab}$ satisfies Left Logical Equivalence and from the definition of $\mathrel{|\!\sim}_{ab}$ (Definition 2.2) it is obvious that Reflexivity and RW holds. It remains to be checked the rules Cut and Cautious Monotony.

Let's suppose that $\alpha \mathrel{|\!\sim}_{ab} \beta$, then the second condition in the rule **E-C-Cut** is satisfied, i.e., $\forall\delta\ [\alpha \rhd \delta \Rightarrow \vdash_\Sigma \delta \to \beta]$. Therefore from **E-C-Cut** and **E-CM** we easily conclude

$$\{\gamma\colon\ \alpha \rhd \gamma\} = \big\{\gamma\colon\ (\alpha \wedge \beta) \rhd \gamma\big\}$$

and hence $C(\alpha \wedge \beta) = C(\alpha)$ (where as usual for a fixed consequence relation $\mathrel{|\!\sim}$ and any formula δ, $C(\delta)$ is the set $\{\theta\colon\ \delta \mathrel{|\!\sim} \theta\}$). That is to say, $\mathrel{|\!\sim}_{ab}$ satisfies Cut and Cautious Monotony. $\square$

Proof of Proposition 2.6. (i) That **RLE** holds is straightforward. To see that **ROR** holds, suppose that $\alpha \rhd (\gamma \vee \rho)$. First note that $(\gamma \vee \rho) \nvdash_\Sigma \neg\gamma$ or $(\gamma \vee \rho) \nvdash_\Sigma \neg\rho$. Otherwise, $(\gamma \vee \rho) \vdash_\Sigma (\neg\gamma \wedge \neg\rho)$ and hence $(\gamma \vee \rho) \vdash_\Sigma \bot$, which is a contradiction since $\rhd$ is an explanatory relation. Therefore by **RA** $\alpha \rhd (\gamma \vee \rho) \wedge \gamma$ or $\alpha \rhd (\gamma \vee \rho) \wedge \rho$. Hence by **RLE** $\alpha \rhd \gamma$ or $\alpha \rhd \rho$.

(ii) and (iii) are straightforward.

(iv) The proof that **RA** implies **E-Disj** is as in (i) above. Conversely suppose that $\rhd$ satisfies **E-Disj** and **RLE**, we want to show that **RA** holds. Let α, γ and γ' be such that $\alpha \rhd \gamma$, $\gamma' \vdash_\Sigma \gamma$ and $\gamma' \nvdash_\Sigma \bot$. Since $\gamma' \vdash_\Sigma \gamma$, we have $\vdash_\Sigma \gamma \leftrightarrow (\gamma' \vee \gamma)$ so by **RLE** $\alpha \rhd (\gamma' \vee \gamma)$. Since by hypothesis $\gamma' \nvdash_\Sigma \bot$ then by **E-Disj** we have $\alpha \rhd \gamma'$.

(v) Suppose $(\alpha_1 \vee \alpha_2) \rhd \gamma$ and $\alpha_i \not\rhd \gamma$ for $i = 1, 2$. We claim that $\gamma \nvdash_\Sigma \alpha_i$ for $i = 1, 2$. Otherwise by **E-CM** we have for some $i \in \{1, 2\}$, $(\alpha_1 \vee \alpha_2) \wedge \alpha_i \rhd \gamma$ and therefore by **LLE** we conclude $\alpha_i \rhd \gamma$ which is a contradiction. Let $\gamma_i = \gamma \wedge \alpha_i$. Since $\gamma \vdash_\Sigma (\alpha_1 \vee \alpha_2)$, then it is clear that γ is equivalent modulo Σ to $\gamma_1 \vee \gamma_2$. On the other hand, $\gamma_i \vdash_\Sigma \alpha_i$ and $\gamma_i \nvdash_\Sigma \bot$ for $i = 1, 2$ (otherwise $\gamma \vdash_\Sigma \alpha_i$ for some i). Finally by **RA** we have that $(\alpha_1 \vee \alpha_2) \rhd \gamma_i$ and by **E-CM** and **LLE** we conclude $\alpha_i \rhd \gamma_i$ for $i = 1, 2$. $\square$

Proof of Theorem 2.8. We already know from Theorem 2.5 that $\mathrel{|\!\sim}_{ab}$ is cumulative, so it remains to be shown that $\mathrel{|\!\sim}_{ab}$ satisfies the rule Or. Let's suppose that $\alpha \mathrel{|\!\sim}_{ab} \rho$ and $\beta \mathrel{|\!\sim}_{ab} \rho$, we will show that $\alpha \vee \beta \mathrel{|\!\sim}_{ab} \rho$. Let γ be such that $(\alpha \vee \beta) \rhd \gamma$, we have to show that $\gamma \vdash_\Sigma \rho$. By Proposition 2.6(v) we have to consider three cases:

(a) $\alpha \rhd \gamma$. Since $\alpha \mathrel{|\!\sim}_{ab} \rho$ then we have $\gamma \vdash_\Sigma \rho$.

(b) $\beta \rhd \gamma$. We conclude that $\gamma \vdash_\Sigma \rho$ as in the first case.

(c) There are γ_1 and γ_2 such that $\vdash_\Sigma \gamma \leftrightarrow (\gamma_1 \vee \gamma_2)$ with $\alpha \rhd \gamma_1$ $\beta \rhd \gamma_2$.

Then, by hypotheses we have $\gamma_i \vdash_\Sigma \rho$ for $i = 1, 2$. Since $\vdash_\Sigma \gamma \leftrightarrow (\gamma_1 \vee \gamma_2)$ we conclude $\gamma \vdash_\Sigma \rho$. $\square$

Proof of Theorem 2.9. By Theorem 2.8 we know that $\mid\!\sim_{ab}$ is preferential. So it remains to be shown that $\mid\!\sim_{ab}$ satisfies **WDR**. We define an auxiliary function F that maps formulas into formulas as follows: $F(\alpha) = \bigvee\{\gamma\colon\ \alpha \rhd \gamma\}$ in case there is γ such that $\alpha \rhd \gamma$, otherwise we let $F(\alpha) = \bot$. Notice that $\alpha \mid\!\sim_{ab} \beta$ iff $F(\alpha) \vdash_\Sigma \beta$. To see that **WDR** holds it clearly suffices to show that $F(\alpha) \wedge F(\beta) \vdash_\Sigma F(\alpha \vee \beta)$. Let $\alpha \rhd \gamma$ and $\beta \rhd \delta$, it is enough to verify that when $\gamma \wedge \delta$ is consistent with Σ, then $(\alpha \vee \beta) \rhd (\gamma \wedge \delta)$. Since $\nvdash_\Sigma \gamma \to \neg\delta$, from **RA** we easily conclude $\alpha \rhd (\gamma \wedge \delta)$ and $\beta \rhd (\gamma \wedge \delta)$, therefore from **LOR** we obtain $(\alpha \vee \beta) \rhd (\gamma \wedge \delta)$. $\Box$

Proof of Proposition 2.10. It is clear that **E-DR** implies **LOR**. To check that **E-DR** implies that $\mid\!\sim_{ab}$ satisfies **DR**, suppose that $\alpha \vee \beta \mid\!\sim_{ab} \rho$ and $\alpha \not\mid\!\sim_{ab} \rho$. We have to show that $\beta \mid\!\sim_{ab} \rho$. Let δ be such that $\beta \rhd \delta$, it suffices to check that $\delta \vdash_\Sigma \rho$. Since $\alpha \not\mid\!\sim_{ab} \rho$, then there is γ such that $\alpha \rhd \gamma$ and $\gamma \nvdash_\Sigma \rho$. By **E-DR** $(\alpha \vee \beta) \rhd \gamma$ or $(\alpha \vee \beta) \rhd \delta$. Since $\alpha \vee \beta \mid\!\sim_{ab} \rho$ and $\gamma \nvdash_\Sigma \rho$, we conclude that $(\alpha \vee \beta) \rhd \delta$. Therefore $\delta \vdash_\Sigma \rho$. $\Box$

Proof of Theorem 2.12. By Theorem 2.8 $\mid\!\sim_{ab}$ is preferential. Thus it suffices to show that $\mid\!\sim_{ab}$ satisfies Rational Monotony. Let α, β and ρ be formulas such that $\alpha \mid\!\sim_{ab} \rho$ and $\alpha \not\mid\!\sim_{ab} \neg\beta$. Let γ be such that $(\alpha \wedge \beta) \rhd \gamma$, we want to show that $\gamma \vdash_\Sigma \rho$. Since $\alpha \not\mid\!\sim_{ab} \neg\beta$, then by definition of $\mid\!\sim_{ab}$ there is δ such that $\alpha \rhd \delta$ and $\delta \nvdash_\Sigma \neg\beta$. By **RA** (see Proposition 2.6(iii)) there is $\delta' \vdash_\Sigma \delta$ such that $\alpha \rhd \delta'$ and $\delta' \vdash_\Sigma \beta$. Therefore by **E-R-Cut** we conclude that $\alpha \rhd \gamma$. Since $\alpha \mid\!\sim_{ab} \rho$, then $\gamma \vdash_\Sigma \rho$. $\Box$

Proof of Proposition 2.13. We will consider three cases.

(Case 1) Suppose that $(\alpha \vee \beta) \mid\!\sim_{ab} \neg\alpha$. In particular, we have that for all $(\alpha \vee \beta) \rhd \gamma$, $\gamma \vdash_\Sigma \beta$. We will show that (b) holds. Let γ be such that $(\alpha \vee \beta) \rhd \gamma$. Then by our hypothesis $\gamma \vdash_\Sigma \beta$. By **E-CM** $(\alpha \vee \beta) \wedge \beta \rhd \gamma$ and by **LLE** $\beta \rhd \gamma$. On the other hand, let γ be such that $\beta \rhd \gamma$, then by **E-CM** $(\alpha \vee \beta) \wedge \beta \rhd \gamma$. Since $(\alpha \vee \beta) \mid\!\sim_{ab} \neg\alpha$, then it follows from **E-C-Cut** that $(\alpha \vee \beta) \rhd \gamma$.

(Case 2) Suppose that $(\alpha \vee \beta) \mid\!\sim_{ab} \neg\beta$. Then as in case 1 it follows that (a) holds.

(Case 3) Suppose that $(\alpha \vee \beta) \not\mid\!\sim_{ab} \neg\alpha$ and $(\alpha \vee \beta) \not\mid\!\sim_{ab} \neg\beta$. We will show that (c) holds. By 2.6(v) it suffices to show that

$$\{\gamma\colon\ \alpha \rhd \gamma\} \cup \{\gamma\colon\ \beta \rhd \gamma\} \subseteq \big\{\gamma\colon\ (\alpha \vee \beta) \rhd \gamma\big\}.$$

By hypothesis there is γ' such that $(\alpha \vee \beta) \rhd \gamma'$ and $\gamma' \nvdash_\Sigma \neg\alpha$. By **RA** we can assume that $\gamma' \vdash_\Sigma \alpha$. Let γ be such that $\alpha \rhd \gamma$, then by **E-CM** $(\alpha \vee \beta) \wedge \alpha \rhd \gamma$. Using γ' and **E-R-Cut** we conclude that $(\alpha \vee \beta) \rhd \gamma$. It can be shown analogously that if $\beta \rhd \gamma$, then $(\alpha \vee \beta) \rhd \gamma$. $\Box$

Proof of Proposition 2.14. Suppose $\alpha \rhd \gamma$ and $\beta \rhd \delta$ and $(\alpha \vee \beta) \not\rhd \delta$, we want to show that $(\alpha \vee \beta) \rhd \gamma$. Since $\vdash \beta \leftrightarrow (\alpha \vee \beta) \wedge \beta$ and $\beta \rhd \delta$ then it follows from **E-R-Cut** that for all γ' if $(\alpha \vee \beta) \rhd \gamma'$, then $\gamma' \nvdash_\Sigma \beta$. Since $\alpha \rhd \gamma$ we have $(\alpha \vee \beta) \wedge \alpha \rhd \gamma$. Suppose, towards a contradiction, that $(\alpha \vee \beta) \not\rhd \gamma$. By **E-C-Cut** there is γ' such that $(\alpha \vee \beta) \rhd \gamma'$ and $\gamma' \nvdash_\Sigma \alpha$. By **RA** (Proposition 2.6) there is γ'' such that $(\alpha \vee \beta) \rhd \gamma''$ and $\gamma'' \vdash_\Sigma \alpha$. Finally, since $\vdash \alpha \leftrightarrow (\alpha \vee \beta) \wedge \alpha$ and $\alpha \rhd \gamma$, then by **E-R-Cut** we conclude that $(\alpha \vee \beta) \rhd \gamma$, a contradiction. $\Box$

Proof of Proposition 2.17. Suppose that $(\alpha \wedge \beta) \rhd \gamma$ and also that $\delta \vdash_\Sigma \beta$ for all δ such that $\alpha \rhd \delta$. It suffices to show that there is δ such that $\alpha \rhd \delta$. Since $(\alpha \wedge \beta) \rhd \gamma$ then (by the definition of an explanatory relation) α is consistent with Σ, therefore by **E-Con$_\Sigma$** there is δ such that $\alpha \rhd \delta$. □

Proof of Proposition 2.20. It is obvious that if $\alpha \vdash_\Sigma \beta$ then $\alpha \mid\!\sim_{ab} \beta$. On the other hand, if $\alpha \nvdash_\Sigma \beta$, then $\alpha \wedge \neg\beta \nvdash_\Sigma \bot$. Thus by **E-Con$_\Sigma$** there is γ such that $(\alpha \wedge \neg\beta) \rhd \gamma$. Therefore by **E-Cut** $\alpha \rhd \gamma$, hence $\alpha \not\mid\!\sim_{ab} \beta$. □

Proof of Proposition 3.3. Let $\tilde{\rhd}$ be the explanatory relation associated with $\mid\!\sim$ and let $\mid\!\sim_{ab}$ be the consequence relation associated with $\tilde{\rhd}$. We will show that $\mid\!\sim$ is equal to $\mid\!\sim_{ab}$. By definition of $\mid\!\sim_{ab}$ and the hypothesis that $\mid\!\sim$ is adequate we have

$$\begin{aligned} C_{ab}(\alpha) &= \bigcap \left\{ Cn(\Sigma \cup \{\gamma\})\colon\ \alpha \,\tilde{\rhd}\, \gamma \right\} \\ &= \bigcap \left\{ Cn(\Sigma \cup \{\gamma\})\colon\ C(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\}) \ \&\ \gamma \nvdash_\Sigma \bot \right\} \\ &= C(\alpha). \end{aligned}$$

Observe that these equalities are valid even in the case that there is no γ such that $\alpha \,\tilde{\rhd}\, \gamma$ (equivalently, when $C(\alpha)$ contains all formulas). □

Proof of Theorem 3.4.

(1) It is obvious from the definition of $\tilde{\rhd}$ that it satisfies **RA**, **E-RW** and **RLE**.

(2) It is obvious that if $\mid\!\sim$ satisfies **LLE**, then $\tilde{\rhd}$ satisfies **LLE**.

(3) Suppose that $\mid\!\sim$ satisfies **Con$_\Sigma$**. It follows easily from the hypothesis that $\mid\!\sim$ is adequate that $\tilde{\rhd}$ satisfies **E-Con$_\Sigma$**.

(4) Suppose that $\mid\!\sim$ satisfies **CM**. To see that $\tilde{\rhd}$ satisfies **E-C-Cut** let us suppose that $(\alpha \wedge \beta) \,\tilde{\rhd}\, \gamma$ and also that $\delta \vdash_\Sigma \beta$ for all δ such that $\alpha \,\tilde{\rhd}\, \delta$. We have to show that $\alpha \,\tilde{\rhd}\, \gamma$. Suppose $\alpha \mid\!\sim \rho$, it suffices to show that $\gamma \vdash_\Sigma \rho$. Since $\mid\!\sim$ is adequate, from the second part of the hypothesis of **E-C-Cut** we conclude that $\alpha \mid\!\sim \beta$. Therefore by **CM** we have $C(\alpha) \subseteq C(\alpha \wedge \beta)$, since $C(\alpha \wedge \beta) \subseteq Cn(\Sigma \cup \{\gamma\})$, then the result follows.

(5) Suppose that $\mid\!\sim$ satisfies the **S**-rule. To see that $\tilde{\rhd}$ satisfies **E-CM** let $\alpha \,\tilde{\rhd}\, \gamma$ and $\gamma \vdash_\Sigma \beta$. We want to show that $(\alpha \wedge \beta) \,\tilde{\rhd}\, \gamma$. Since γ is consistent with Σ, it suffices to show that $C(\alpha \wedge \beta) \subseteq Cn(\Sigma \cup \{\gamma\})$. Let $\alpha \wedge \beta \mid\!\sim \rho$, then by the **S**-rule $\alpha \mid\!\sim \beta \to \rho$. Since $\alpha \,\tilde{\rhd}\, \gamma$, then $\gamma \vdash_\Sigma \beta \to \rho$. Hence $\gamma \vdash_\Sigma \rho$.

(6) Suppose $\mid\!\sim$ satisfies **WDR**. We will show that $\tilde{\rhd}$ satisfies **LOR**. Suppose $\alpha \,\tilde{\rhd}\, \gamma$ and $\beta \,\tilde{\rhd}\, \gamma$. By **WDR** we have that $C(\alpha \vee \beta) \subseteq Cn(C(\alpha) \cup C(\beta))$. Then it is clear that $(\alpha \vee \beta) \,\tilde{\rhd}\, \gamma$.

(7) Suppose $\mid\!\sim$ is preferential and satisfies **DR**. We will show that $\tilde{\rhd}$ satisfies **E-DR**. Suppose $\alpha \,\tilde{\rhd}\, \gamma$, $\beta \,\tilde{\rhd}\, \rho$ and $(\alpha \vee \beta) \,\tilde{\not\rhd}\, \gamma$. Then there is δ such that $\alpha \vee \beta \mid\!\sim \delta$ and $\gamma \nvdash_\Sigma \delta$. Since $\alpha \,\tilde{\rhd}\, \gamma$ and $C(\alpha \vee \beta) \subseteq C(\alpha) \cup C(\beta)$ we have $\delta \in C(\beta)$. Now consider any δ' such that $\delta' \in C(\alpha \vee \beta)$. We want to show that $\rho \vdash_\Sigma \delta'$. By preferentiality $\delta \wedge \delta' \in C(\alpha \vee \beta)$. But $\delta \wedge \delta' \notin C(\alpha)$, otherwise $\gamma \vdash_\Sigma \delta \wedge \delta'$ and therefore $\gamma \vdash_\Sigma \delta$ which is a contradiction. Then by **DR** $\delta \wedge \delta' \in C(\beta)$. Hence $\rho \vdash_\Sigma \delta \wedge \delta'$ and thus $\rho \vdash_\Sigma \delta'$.

(8) Suppose $\mid\!\sim$ satisfies **RM**. We will show that $\tilde{\rhd}$ satisfies **E-R-Cut**. Suppose $(\alpha \wedge \beta)\,\tilde{\rhd}\,\gamma$ and there is δ such that $\alpha\,\tilde{\rhd}\,\delta$ with $\delta \vdash_\Sigma \beta$. From the last assumption and the definition of $\tilde{\rhd}$ we conclude that $\alpha \not\mid\!\sim \neg\beta$. Therefore by **RM** we have $C(\alpha) \subseteq C(\alpha \wedge \beta)$, and the result follows.

(9) Suppose that $\mid\!\sim$ is monotone. Since $\mid\!\sim$ is monotone, then $C(\alpha) \subseteq C(\alpha \wedge \beta)$. Therefore, if $(\alpha \wedge \beta)\,\tilde{\rhd}\,\gamma$ then $\alpha\,\tilde{\rhd}\,\gamma$. This says that $\tilde{\rhd}$ satisfies **E-Cut**. $\square$

Proof of Proposition 3.5. For every α, let $F(\alpha)$ be a formula such that $C(\alpha) = Cn(F(\alpha))$. Let us define $\rhd$ as follows: $\alpha \rhd \gamma$ if $\gamma \nvdash_\Sigma \bot$ and $\gamma \equiv F(\alpha)$. It is obvious that $\rhd$ is indeed an explanatory relation satisfying **RLE**. Let $\mid\!\sim_{ab}$ be the consequence relation associate with $\rhd$. It is easy to see that $\mid\!\sim$ is equal to $\mid\!\sim_{ab}$. Now we will check the other postulates. It follows that **LLE** (for $\rhd$) follows from **LLE** for $\mid\!\sim$. To see that **E-CM** holds, suppose $\alpha \rhd \gamma$ and $\gamma \vdash_\Sigma \beta$. We need to show that $(\alpha \wedge \beta) \rhd \gamma$. By hypothesis $F(\alpha) \vdash_\Sigma \beta$, then it follows that $\alpha \mid\!\sim \beta$. Since $\mid\!\sim$ is cumulative, then $C(\alpha) = C(\alpha \wedge \beta)$. From this it follows that $F(\alpha) \equiv F(\alpha \wedge \beta)$ and therefore $(\alpha \wedge \beta) \rhd \gamma$. The proof that **E-C-Cut** holds is similar. $\square$

Proof of Proposition 3.8. (i) $\Rightarrow$ (ii). It is obvious that any causal relation satisfies **RA**. To check that **C** holds let α and γ be two formulas consistent with Σ. Suppose that for all δ consistent with Σ such that $\delta \vdash_\Sigma \gamma$ there is ρ such that $\rho \vdash_\Sigma \delta$ and $\alpha \rhd \rho$. We want to show that $C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})$. Let $\alpha \mid\!\sim_{ab} \beta$ and suppose toward a contradiction that $\gamma \nvdash_\Sigma \beta$. Since $\gamma \wedge \neg\beta$ is consistent with Σ then by the hypothesis in **C** there is ρ such that $\alpha \rhd \rho$ and $\rho \vdash_\Sigma \gamma \wedge \neg\beta$. On the other hand, since $\alpha \mid\!\sim_{ab} \beta$, then $\rho \vdash_\Sigma \beta$. Thus $\rho \vdash_\Sigma \bot$ which is a contradiction.

(ii) $\Rightarrow$ (i). Suppose that $\rhd$ satisfies **RA** and **C**. It suffices to show that if $\gamma \nvdash_\Sigma \bot$ and $C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})$, then $\alpha \rhd \gamma$. Let δ be any formula consistent with Σ such that $\delta \vdash_\Sigma \gamma$. Then there must exist ρ such that $\alpha \rhd \rho$ and $\rho \nvdash_\Sigma \neg\delta$ (otherwise $\neg\delta \in C_{ab}(\alpha)$ which is not possible). Then $\rho \wedge \delta$ is consistent with Σ. By **RA** we conclude that $\alpha \rhd \rho \wedge \delta$. Therefore by **C** we get that $\alpha \rhd \gamma$. $\square$

Proof of Proposition 3.9. (i) implies (ii) follows from the fact that when **E-Con$_\Sigma$** and **E-Cut** hold, then $\mid\!\sim_{ab} = \vdash_\Sigma$ (see Proposition 2.20). Therefore, since $\rhd$ is causal, then $\alpha \rhd \gamma$ iff $Cn(\Sigma \cup \{\alpha\}) \subseteq Cn(\Sigma \cup \{\gamma\})$ iff $\gamma \vdash_\Sigma \alpha$. On the other hand, to see that (ii) implies (i), just notice that an explanatory relation defined as in (ii) satisfies **E-Cut**, **E-Con$_\Sigma$** and it is causal. $\square$

Proof of Proposition 3.12. It suffices to show that $\rhd$ satisfies **C**. Let $\gamma_1, \ldots, \gamma_k$ be an upper bound for $\{\gamma\colon \alpha \rhd \gamma\}$. Let $\theta = \gamma_1 \vee \cdots \vee \gamma_k$. By **E-RW** we have that $\alpha \rhd \theta$. Let α and γ be formulas consistent with Σ. Suppose that for all δ such that $\delta \nvdash_\Sigma \bot$ and $\delta \vdash_\Sigma \gamma$ there is ρ such that $\alpha \rhd \rho$ and $\rho \vdash_\Sigma \delta$. We want to show that $\alpha \rhd \gamma$. Suppose that $\alpha \not\rhd \gamma$ towards a contradiction. Then by **RA** we have that $\gamma \nvdash_\Sigma \theta$. Therefore $\gamma \wedge \neg\theta$ is consistent with Σ. By hypothesis there is ρ such that $\alpha \rhd \rho$ and $\rho \vdash_\Sigma \gamma \wedge \neg\theta$, which contradicts that θ is an upper bound. $\square$

Proof of Proposition 4.3. Since $*$ preserves consistency, then it is clear that $\mid\!<$ is a weak explanatory relation (as defined in Section 4.2). It is obvious that **LLE**, **RLE**, **E-Cut** and full reflexivity holds. Notice that **LOR** follows from **E-Cut** and **LLE**. To check **E-CM**, assume that $\alpha |\!< \gamma$ and also that $\gamma \vdash \beta$. Then $\gamma, \alpha \in K * \gamma$. Thus $\alpha \wedge \beta \in K * \gamma$. Finally, **E-RW** follows from the Or rule for $|\!\sim_K$ and **ROR** follows from **DR** for $|\!\sim_K$. $\square$

Proof of Proposition 4.4. From $\mathbf{Con}_\Sigma$ it follows that $|\!<$ is a weak explanatory relation.

(i) It is clear that $|\!<$ is transitive, reflexive for Σ-consistent formulas and satisfies **LLE**, **RLE**. **E-RW** follows easily from the Or rule. To check **E-CM**, let us assume that $\alpha |\!< \gamma$ and $\gamma \vdash_\Sigma \beta$. Let $\alpha \wedge \beta |\!\sim \rho$, then by the **S**-rule we have that $\alpha |\!\sim \beta \to \rho$. By hypothesis $C(\alpha) \subseteq C(\gamma)$, thus $\gamma |\!\sim \beta \to \rho$. By preferentiality and $\mathbf{Con}_\Sigma$ from $\gamma \vdash_\Sigma \beta$ is easy to obtain $\gamma |\!\sim \beta$. Hence by **RW** $\gamma |\!\sim \rho$. Therefore $C(\alpha \wedge \beta) \subseteq C(\gamma)$. To check **E-C-Cut**, assume that $\alpha \wedge \beta |\!< \gamma$ and also that $\delta \vdash_\Sigma \beta$ for all δ such that $\alpha |\!< \delta$. In particular, since $|\!<$ is reflexive, we have that $\alpha \vdash_\Sigma \beta$. Thus $\alpha |\!\sim \beta$ and therefore $C(\alpha) = C(\alpha \wedge \beta)$.

(ii) **DR** says that $C(\alpha \vee \beta) \subseteq C(\alpha) \cup C(\beta)$ from which it is obvious that **LOR** holds.

(iii) Suppose that $|\!\sim$ is rational. We will use the following well known fact about rational relations. For every pair of formulas α and β one of the following holds:

(i) $C(\alpha \vee \beta) = C(\alpha)$,
(ii) $C(\alpha \vee \beta) = C(\beta)$,
(iii) $C(\alpha \vee \beta) = C(\alpha) \cap C(\beta)$.

From this is obvious that **E-DR** and **ROR** hold. It remains to be checked that **E-R-Cut** holds. Suppose that $\alpha \wedge \beta |\!< \gamma$ and also that there is δ such that $\alpha |\!< \delta$ and $\delta \vdash_\Sigma \beta$. By **RM** it suffices to show that $\alpha \not|\!\sim \neg\beta$. Assume $\alpha |\!\sim \neg\beta$ towards a contradiction. Since $\alpha |\!< \delta$, then $\delta |\!\sim \neg\beta$. Since $\delta \vdash_\Sigma \beta$, then by preferentiality and $\mathbf{Con}_\Sigma$, $\delta |\!\sim \beta$. Therefore $\delta |\!\sim \bot$ which contradicts the fact that δ is Σ-consistent. $\square$

References

[1] C.E. Alchourrón, P. Gärdenfors, D. Makinson, On the logic of theory change, J. Symbolic Logic 50 (1985) 510–530.
[2] A. Aliseda-Llera, Seeking explanations: Abduction in logic, philosophy of science and artificial intelligence, Ph.D. Thesis, Stanford University, 1997. http://www.wins.uva.nl/research/illc/wwwdissertations.html.
[3] C. Boutilier, V. Becher, Abduction as belief revision, Artificial Intelligence 77 (1995) 43–94.
[4] M. Cialdea Mayer, F. Pirri, Abduction is not deduction-in-reverse, J. IGPL 4 (1) (1996) 1–14.
[5] P.A. Flach, Rationality postulates for induction, in: Y. Shoham (Ed.), Proc. Sixth Conference of Theoretical Aspects of Rationality and Knowledge (TARK-96), 1996, pp. 267–281.
[6] M. Freund, Injective models and disjunctive relations, J. Logic Comput. 3 (1993) 231–247.
[7] P. Gärdenfors, Knowledge in Flux: Modeling the Dynamics of Epistemic States, MIT Press, Cambridge, MA, 1988.
[8] P. Gärdenfors, D. Makinson, Relations between the logic of theory change and nonmonotonic logic, in: The Logic of Theory Change, Workshop, Konstanz, FRG, Lecture Notes in Artificial Intelligence, Vol. 465, Springer, Berlin, 1989, pp. 185–205.
[9] P. Gärdenfors, D. Makinson, Nonmonotonic inferences based on expectations, Artificial Intelligence 65 (1994) 197–245.
[10] S. Kraus, D. Lehmann, M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, Artificial Intelligence 44 (1) (1990) 167–207.
[11] D. Lehmann, Draft on the logic of induction, Technical Report, Institute of Computer Science, Hebrew University, 1995.
[12] D. Lehmann, M. Magidor, What does a conditional knowledge base entail?, Artificial Intelligence 55 (1992) 1–60.

[13] H.J. Levesque, A knowledge level account of abduction, in: Proc. IJCAI-89, Detroit, MI, 1989, pp. 1061–1067.
[14] J. Lobo, C. Uzcátegui, Abductive change operators, Fundamenta Informaticae 27 (4) (1996) 385–411.
[15] J. Lobo, C. Uzcátegui, Abductive consequence relations, Artificial Intelligence 89 (1–2) (1997) 149–171.
[16] D. Makinson, General patterns in nonmonotonic reasoning, in: C. Hogger D. Gabbay, J. Robinson (Eds.), Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. III, Nonmonotonic Reasoning and Uncertain Reasoning, Oxford University Press, Oxford, 1994.
[17] G. Paul, Approaches to abductive reasoning: An overview, Artificial Intelligence Review 7 (1993) 109–152.
[18] R. Pino-Pérez, C. Uzcátegui, On representation theorems for nonmonotonic consequence relations, J. Symbolic Logic, to appear.
[19] R. Pino-Pérez, C. Uzcátegui, Ordering explanations, in preparation.
[20] W. Zadrozny, On the rules of abduction, Ann. Math. Artificial Intelligence 9 (1993) 387–419.