# Preferences and Explanations 

Ramón Pino-Pérez and Carlos Uzcátegui<br>Departamento de Matemáticas<br>Facultad de Ciencias<br>Universidad de Los Andes<br>Mérida 5101, Venezuela<br>pino@ciens.ula.ve and uzca@ciens.ula.ve

February 6, 2003


#### Abstract

In abductive reasoning, preference criteria for selecting the best explanation are regarded as qualitative properties (like being simpler or more plausible) which are beyond the pure causal or deductive relationship between an explanandum and its explanations. This paper is a contribution to the clarification of the relationship between preference criteria and structural properties of explanatory reasoning. We present a detailed analysis of the connection between the logical properties satisfied by a logic-based explanatory process and the structural properties satisfied by the criterion used for selecting the preferred explanations. Namely, we characterize the postulates introduced in a previous work (AIJ 111(2), 1999) as those satisfied by explanatory relations defined by preference relations over formulas. Several examples illustrating our results are analyzed, including well known preference criteria like expectation orders, preferential orders and other selection criteria that have appeared in the literature.


Keywords: Abduction; explanatory reasoning; preference criteria for selecting explanations.

## 1 Introduction

Abduction is usually defined as the process of inferring the best explanation of an observation. The emphasis placed on preferred explanations rather than plain explanations is perhaps the most distinct feature of explanatory reasoning. Preference criteria for selecting the best explanation are regarded as qualitative properties (like being simpler or more plausible) which are beyond the pure causal or deductive relationship between an explanandum and its explanations. Thus preference criteria are treated as an external device which works on top of the logical or deductive part of the explanatory mechanism. However, the relationship between preference criteria and structural properties of explanatory reasoning has not been clearly delineated. This paper is a contribution to the clarification of this issue. We refer the reader to the books $[10,14]$ for an updated overall picture of abductive reasoning.

In the logic-based approach to abduction, there is a very natural notion of a (possible or plain) explanation. The background theory is a consistent set of formulas $\Sigma$ in a propositional language which describes the causal relations in the domain of application. A formula $\gamma$ (consistent with $\Sigma$ ) is said to be an explanation of $\alpha$, if $\Sigma \cup\{\gamma\} \vdash \alpha$. The set of explanations of $\alpha$ will be denoted by

Expla $(\alpha)$. As we said, explanatory reasoning is concerned with preferred explanations rather than just plain explanations. So, explaining an observation $\alpha$ requires that some formulas in $\operatorname{Expl}(\alpha)$ must be "selected" as preferred explanations. Let us see two paradigmatic examples of preference criteria.

A standard method to assign a degree of plausibility (expectation, simplicity, etc.) to a formula is through a strict partial order $\prec$ over the set of formulas. It is customary to write $\gamma \prec \delta$ when $\gamma$ has a strictly larger degree (of acceptance) than $\delta$. Thus for a given a formula $\alpha$, the preferred explanations of $\alpha$ are commonly defined as the $\prec$-minimal elements of $\operatorname{Expla}(\alpha)$ [6, 20]. The order $\prec$ can be interpreted in many disparate ways. Well known examples are the expectation orders of Gärdenfors and Makinson [15, 16], the possibilistic orders of Dubois and Prade [7] and the preferential orders of Freund [13]. All of them were developed to measure the "plausibility" or "degree of expectation" of a formula. They are, generally speaking, extensions or variations of the (reversed) lattice order on formulas which says that logically weak formulas are preferred (i.e. $\gamma \preceq \delta$ if $\delta \vdash \gamma$ ). On the other hand, there are also orders which are tied more closely to the syntax of formulas. A typical example is the simplicity criterion used by Levesque [20] which says that a formula $\gamma$ is simpler than a formula $\delta$ if the set of literals appearing in $\gamma$ is a proper subset of the set of literals in $\delta$, but keeping track of their polarity. For example $(p \wedge \neg q)$ is simpler than $(\neg q \vee p \vee r)$ but not than $(q \vee p \vee r)$. All these examples will be analyzed in this paper.

Belief revision, abduction and nonmonotonic reasoning share some common grounds (see [2, 5, $8,16,21,22]$ ). A typical example illustrating such a connection is the following preference criteria based on the possible world semantics. For each formula $\alpha$ we denote by $\bmod (\alpha)$ the models of $\alpha$ (i.e. those valuations of the language which assign 1 to $\alpha$ ). The models of a set of formulas $\Sigma$ are defined analogously and denoted by $\bmod (\Sigma)$. A measure of the relative plausibility between models is given by a strict partial order $\sqsubset$ over $\bmod (\Sigma)$, where $M \sqsubset N$ is read as $M$ is more plausible than $N$. For each formula $\alpha$ we define $\min (\alpha)$ as the $\sqsubset$-minimal elements of $\bmod (\Sigma \cup\{\alpha\})$. The collection $\min (\alpha)$ contains the more plausible (normal, most expected, etc.) models of $\alpha$. A standard inference process associated to $\sqsubset$ is defined as follows [18]: a formula $\alpha$ is said to (normally) entail $\beta$ (denoted $\alpha{\nu_{\sqsubset}} \beta$ ) if $M \models \beta$ for all $M \in \min (\alpha)$ (i.e. $\min (\alpha) \subseteq \bmod (\beta)$ ). The relation $\mu_{\llcorner }$is a (nonmonotonic) consequence relation. Thus the normal or most expected consequences of a formula $\alpha$ are determined by $\min (\alpha)$. From this point of view, it is natural to use $\min (\alpha)$ as a tool to select some explanations of (an observation) $\alpha$. For example, we could say that $\gamma$ is a preferred explanation of $\alpha$ if $\bmod (\Sigma \cup\{\gamma\}) \subseteq \min (\alpha)$. In other words, $\gamma$ is a preferred explanation of $\alpha$ because it (classically) implies everything that $\alpha$ normally implies. We will analyze several such methods in $\S 2$.

But what do all these criteria have in common, and how rational is "inferring to the best explanation" based on any of them? The second example above suggests that such questions can be naturally approached with the well known KLM methodology developed by Kraus, Lehmann and Magidor to study nonmonotonic reasoning [18, 19, 24]. Based on this intuition, a list of structural properties for explanatory reasoning was introduced in [26] (see also [1, 6, 9, 11, 12]). Examples of such properties are the following:
(a) Explanatory Cautious Monotony: if $\gamma$ is a preferred explanation of $\alpha$ and $\Sigma \cup\{\gamma\} \vdash \beta$, then $\gamma$ is also a preferred explanation of $\alpha \wedge \beta$.
(b) Explanatory Cautious Cut: if $\gamma$ is a preferred explanation of $\alpha \wedge \beta$ but it is not a preferred explanation of $\alpha$, then there must exist a preferred explanation $\delta$ of $\alpha$ such that $\Sigma \cup\{\delta\} \nvdash \beta$.
(c) Right Strengthening: if $\gamma$ is a preferred explanation of $\alpha$ and $\gamma^{\prime}$ is a formula consistent with $\Sigma$ such that $\gamma^{\prime} \vdash \gamma$, then $\gamma^{\prime}$ is also a preferred explanation of $\alpha$.
(d) Right Or: if $\gamma$ and $\delta$ are two preferred explanations of $\alpha$, then $\gamma \vee \delta$ is also a preferred explanation of $\alpha$.
(e) Left Or: if $\gamma$ is a preferred explanation of $\alpha$ and also of $\beta$, then $\gamma$ is a preferred explanation of $\alpha \vee \beta$.

These rules (together with others that will be introduced later) can be used to classify explanatory relations in terms of their logical structure. The intuition behind this classification is briefly as follows (more details will be given in §2). We have argued in [26] that the common (classical) consequences of all preferred explanations of an observation could be as relevant as the explanations themselves. These common consequences provide useful elements to conjecture what else to expect, or what else should be present, when a particular fact is observed ${ }^{1}$. This map, from an observation to its abductive consequences, is a well-defined deductive operation. The analysis of this operation through the KLM methodology gave the initial intuition for the postulates.

The rules will be applied in the following general setting. We will consider binary relations over formulas generically denoted with the symbol $\triangleright$. The only requirement is that whenever $\alpha \triangleright \gamma$, then $\Sigma \cup\{\gamma\} \vdash \alpha$. They are called explanatory relations and $\alpha \triangleright \gamma$ is read as " $\gamma$ is a preferred explanation of $\alpha$ ". The examples above can be presented in this setting as follows

$$
\begin{equation*}
\alpha \triangleright_{\prec} \gamma \stackrel{\text { def }}{\Leftrightarrow} \gamma \in \min (\operatorname{Expla}(\alpha), \prec) \tag{1}
\end{equation*}
$$

and

$$
\begin{equation*}
\alpha \triangleright_{\sqsubset} \gamma \stackrel{\text { def }}{\Leftrightarrow} \bmod (\Sigma \cup\{\gamma\}) \subseteq \min (\alpha) \tag{2}
\end{equation*}
$$

For instance, $\triangleright_{\sqsubset}$ always satisfies all five postulates stated above, regardless of $\sqsubset$. On the other hand, $\nabla_{\prec}$ always satisfies (a) and (b) but, in general, does not satisfy neither (c), (d) nor (e).

The main topic of this paper is to study the relationship between the structural properties for explanatory relations and the properties of the preference relation used to select explanations. On the one hand, we will determine which structural properties of $\prec$ (like being a total order, or more general, filtered, etc) are sufficient in order that its associated explanatory relation $\nabla_{<}$satisfies a given postulate. An example of this type of results, already mentioned above, is that (for a finite language) $\nabla_{<}$satisfies Explanatory Cautious Monotony and Explanatory Cautious Cut for any partial order $\prec$. On the other hand, we will show that if a given explanatory relation $\triangleright$ satisfies certain basic postulates, then there is a preference relation $\prec$ such that $\triangleright=\triangleright_{\mathcal{L}}$. In other words, we will show that, under fairly general conditions, every explanatory relation is of the form (1). For example, for a finite language this happens when Explanatory Cautious Monotony, Explanatory Cautious Cut and Left Or are satisfied. In particular, for a given order $\sqsubset$ on $\bmod (\Sigma)$ there is a preference relation $\prec$ on formulas such that $\triangleright_{\sqsubset}$ is equal to $\nabla_{\swarrow}$.

It is well known that many forms of defeasible reasoning are based on some notion of minimality ${ }^{2}$. Moreover, several systems of postulates developed to study nonmonotonic reasoning are

[^0]characterized in terms of orders over formulas [7, 13, 15]. Thus our results are very natural in the light of these facts. But in addition, they provide evidence about the correctness of the proposed postulates for explanatory reasoning ${ }^{3}$. We will come back to this point in the concluding remarks.

This paper is organized as follows. In section 2 we present a variety of examples of explanatory relations. They will illustrate the nature of the postulates. The analysis of theses examples will help in understanding the scope of the results presented and thus it is an important part of the paper. In section 3 we recall the main structural rules for explanatory reasoning introduced in [26] and a basic classification of explanatory relations based on them. In section 4 we present the structural properties to classify preference relations. In section 5 we define the essential preference relation associated to an explanatory relation and state the main results of the paper. In section 6 we revisit the examples to see how our results can be applied. A summary of the results and some final comments are given in section 7. In appendix A we give the proofs of the main results.

A first version of some of the results presented in this paper appeared in [28].
Acknowledgments: We would like to thank the anonymous referee for his comments and criticism which have improved the quality of the paper. This work was partially supported by the cooperation program \#PI98003494 between CNRS (France) and CONICIT (Venezuela).

## 2 Explanatory relations

The background theory describing the causal relations of the domain of application will be a consistent set of formulas in a classical propositional language. It will be denoted by $\Sigma$. The classical consequence relation is denoted by $\vdash$. For a set of formulas $\Gamma$ we write $\alpha \vdash_{\Gamma} \beta$ when $\Gamma \cup\{\alpha\} \vdash \beta$. $C n(\Gamma)$ will denote the set of all classical consequences of $\Gamma$.

We now introduce the notion of an explanation of a formula with respect to $\Sigma$.
Definition 2.1. For every formula $\alpha$, the collection of explanations of $\alpha$ with respect to $\Sigma$ is denoted by Expla( $\alpha$ ) and is defined as follows:

$$
\operatorname{Expla}(\alpha)=\left\{\gamma: \gamma \vdash_{\Sigma} \perp \& \gamma \vdash_{\Sigma} \alpha\right\}
$$

We have ruled out trivial explanations by asking that $\gamma$ has to be consistent with $\Sigma$. Now we formally define the main object of study of this paper.

Definition 2.2. Let $\Sigma$ be a background theory. An explanatory relation for $\Sigma$ will be any binary relation $\triangleright$ such that for every $\alpha$ and $\gamma$,

$$
\alpha \triangleright \gamma \Rightarrow \gamma \nvdash_{\Sigma} \perp \text { and } \gamma \vdash_{\Sigma} \alpha
$$

We read $\alpha \triangleright \gamma$ as $\gamma$ is a preferred explanation (with respect to $\Sigma$ ) of $\alpha$.
In explanatory reasoning the input is an observation and the output is an explanation, that is the reason to write $\alpha \triangleright \gamma$ with $\alpha$ as input and $\gamma$ as output.

We will present several examples illustrating a variety of methods to select explanations and, therefore, to define explanatory relations.

[^1]
### 2.1 Cautious explanations

It is often claimed that logically weak explanations are preferred. The following interesting variation of this idea is taken from [6]. Let $\Pi$ be a collection of formulas and define $\gamma \prec_{\Pi} \delta$ if $\delta \vdash_{\Pi} \gamma$ and $\gamma \nvdash_{\Pi} \delta$ (recall that $\alpha \vdash_{\Pi} \beta$ means $\Pi \vdash \alpha \rightarrow \beta$ ). Let $\triangleright_{\Pi}$ be the explanatory relation given by (1) with $\prec=\prec_{\Pi}$.

When $\Pi=\Sigma$ we have a very simple notion of preferred explanation, since $\gamma \triangleright_{\Sigma} \alpha$ if and only if $\gamma$ is equivalent to $\alpha$ modulo $\Sigma$. The more interesting cases occur when $\Pi$ is a proper extension of $\Sigma$. Assuming $\Sigma \subseteq \Pi$, we have the following

$$
\begin{equation*}
\alpha \triangleright_{\Pi} \gamma \Leftrightarrow \gamma \vdash_{\Sigma} \perp \& \gamma \vdash_{\Sigma} \alpha \& \alpha \vdash_{\Pi} \gamma \tag{3}
\end{equation*}
$$

From this it follows, for instance, that $\triangleright_{\Pi}$ satisfies the rule Left Or stated in the introduction.
For a given (plain) explanation $\gamma$ of $\alpha$ we have now a simple method to establish a preference criterion with respect to which $\gamma$ is a preferred explanation. In fact, suppose $\gamma \in \operatorname{Expla}(\alpha)$ and let $\Pi=\Sigma \cup\{\alpha \rightarrow \gamma\}$ (of course, the interesting case is when $\alpha \nvdash_{\Sigma} \gamma$ ), then from the equivalence above it follows that $\alpha \triangleright_{\Pi} \gamma$.

The following concrete example is taken from [17], where such an extension $\Pi$ of $\Sigma$ was used to close or complete a causal theory. Consider the following collection of formulas

$$
\Sigma=\left\{\begin{array}{ccc}
s & \rightarrow & l \\
r & \rightarrow & l \\
r & \rightarrow & d \\
n \wedge w & \rightarrow & s
\end{array}\right.
$$

Let us suppose that we are only interested in explaining two formulas $l$ and $d$ (called effects in [17]) either of which will be denoted by $O$. Also suppose that the set $C$ of possible causes consists of $s$, $r, n$ and $w$. An abductive explanation of $O$ is an explanation of $O$ which is a conjunction of atoms from $C$ and in addition it is required to be of minimal size. To define the extension $\Pi$, we consider $\gamma(O)$ the disjunction of all abductive explanations of $O$ (the so called cautious explanation) and add to $\Sigma$ the formula $O \rightarrow \gamma(O)$. The cautious explanation of $l$ is equivalent (modulo $\Sigma$ ) to $r \vee s$ and the cautious explanation of $d$ is $r$. So we have the following extension

$$
\Pi=\Sigma \cup\{l \rightarrow(r \vee s)\} \cup\{d \rightarrow r\}
$$

Then $l \triangleright_{\Pi}(r \vee s)$ and $d \triangleright_{\Pi} r$. Notice that $w \wedge n, s, r$ are not preferred explanations of $l$ with respect to $\prec_{\Pi}\left(\right.$ since $(r \vee s) \prec_{\Pi}(w \wedge n),(r \vee s) \prec_{\Pi} s$ and $\left.(r \vee s) \prec_{\Pi} r\right)$. The same result was obtained by the explanatory closure defined in [17]. However, there is an important difference. We are not deducing the cautious explanation $\gamma(O)$ from $\Pi$ together with a given effect $O$, as in [17]. But rather, based on the preference relation $\prec_{\Pi}$, we are inferring $\gamma(O)$ from $\Sigma$ together with $O$.

### 2.2 Simple explanations

Many accounts of explanatory reasoning argue in favor of minimizing the number of literals present in an explanation as a simplicity criterion. We will present two examples of this type.

Let us consider the syntactic criterion of simplicity used by Levesque [20], already mentioned in the introduction. The precise definition is as follows. The set of literals in a formula $\alpha$ is denoted
by $\mathrm{L}(\alpha)$ and defined recursively by $\mathrm{L}(\perp)=\emptyset ; \mathrm{L}(p)=\{p\}$ for any atom $p, \mathrm{~L}(\neg \alpha)=\{\bar{m}: m \in \mathrm{~L}(\alpha)\}$ and $\mathrm{L}(\alpha \vee \beta)=\mathrm{L}(\alpha \wedge \beta)=\mathrm{L}(\alpha) \cup \mathrm{L}(\beta)$. Let $\prec_{\text {lit }}$ be given by

$$
\gamma \prec_{l i t} \delta \stackrel{\text { def }}{\Leftrightarrow} \mathrm{L}(\gamma) \nsubseteq \mathrm{L}(\delta)
$$

The order $\prec_{l i t}$ is extremely sensitive to the syntax, for instance, $\neg a \prec_{l i t} a \wedge(\neg a \vee b)$ but $\neg a \nprec{ }_{l i t} a \wedge b$. Let $\triangleright_{l i t}$ denote the explanatory relation associated to $\prec_{l i t}$ given by (1).

In case we are interested only in the number of literals appearing in a formula as a simplicity criterion we can modify the previous order as follows. Let $|\gamma|$ denote the cardinality of $L(\gamma)$ and define $\prec_{p}$ by

$$
\gamma \prec_{p} \delta \stackrel{\text { def }}{\Leftrightarrow}|\gamma|<|\delta|
$$

Let $\triangleright_{p}$ be the explanatory relation given by $\prec_{p}$. It is clear that if $\alpha \triangleright_{p} \gamma$, then $\alpha \triangleright_{\text {lit }} \gamma$, but the converse is not true.

The explanatory relations $\nabla_{l i t}, \triangleright_{p}$ and $\triangleright_{\Pi}$ are structurally different. For instance, we have mentioned already that Left Or holds for $\triangleright_{\Pi}$ but this is not the case for the other two. To see this, suppose $\Sigma$ is empty and let $\alpha=(a \wedge b) \vee(a \wedge c)$ and $\beta=(a \wedge \neg b) \vee(a \wedge c)$. Notice that $\alpha \vee \beta$ is equivalent to $a$. Then $\alpha \triangleright_{p} a \wedge c, \beta \triangleright_{p} a \wedge c$. But $(\alpha \vee \beta) \not{ }_{p} a \wedge c$ because $|a|<|a \wedge c|$. The same example works for $\triangleright_{l i t}$.

The five postulates stated in the introduction are not enough to distinguish $\triangleright_{p}$ from $\triangleright_{l i t}$, since both relations only satisfy Explanatory Cautious Monotony and Explanatory Cautious Cut. We will introduce later another postulate showing that they are indeed structurally different.

We can slightly modify the relation $\prec_{p}$ by simply restricting it to (formulas equivalent to a) conjunction of literals (thus putting above them every formula which is not equivalent to a conjunction of literals). The point is that under this modified preference criterion, the preferred explanations of a formula are exactly their prime implicants. Prime implicants have been widely used for modelling explanatory reasoning and diagnosis (for a general discussion about it see, for instance, $[25, \S 1]$ and references therein).

### 2.3 Some explanatory relations based on minimal models

Preferential models are frequently used to define structures for modelling knowledge representation problems [18]. As it was mentioned in the introduction, they can also be used to construct explanatory relations. For the sake of simplicity, we will assume that the language is finite ${ }^{4}$. Fix a strict partial order $\sqsubset$ on $\bmod (\Sigma)$ and define the notion of a minimal model of a formula $\alpha$ as follows

$$
\min (\alpha)=\{N: N \models \Sigma \cup\{\alpha\} \& M \not \models \alpha \text { for all } M \sqsubset N\}
$$

The associated consequence relation $\mathcal{N}_{\sqsubset}$ is defined by

$$
\begin{equation*}
\alpha \psi_{\sqsubset} \beta \text { iff } \min (\alpha) \subseteq \bmod (\beta) \tag{4}
\end{equation*}
$$

If there is no confusion about which order $\sqsubset$ is used we will just write $\sim$. Let $C(\alpha)=\{\beta: \alpha \sim \beta\}$. The relation $\sim$ is preferential in the sense of [18].

[^2]As we argued in the introduction, to select explanations of a formula $\alpha$ it is natural to look only at $\min (\alpha)$ instead of $\bmod (\alpha)$. There are several ways of doing this. We will present three of them.

The first one is called causal explanatory relation. It is denoted by $\triangleright_{c}$ and defined as follows

$$
\begin{equation*}
\alpha \triangleright_{c} \gamma \stackrel{\text { def }}{\Leftrightarrow} \bmod (\Sigma \cup\{\gamma\}) \subseteq \min (\alpha) \tag{5}
\end{equation*}
$$

for any pair of consistent (with $\Sigma$ ) formulas $\alpha$ and $\gamma$. In the introduction we denoted it by $\triangleright_{\square}$. It is easy to verify that $\alpha \triangleright_{c} \gamma$ if and only if $C(\alpha) \subseteq C n(\Sigma \cup\{\gamma\})$, where $\sim$ is given by (4). The most characteristic feature of $\triangleright_{c}$ is that it satisfies Right Strengthening and Right Or. In fact, any explanatory relation satisfying these two postulates has to be (essentially) of the form $\triangleright_{c}$ [26, Proposition 3.8].

The second example is called Strong Epistemic explanation.

$$
\begin{equation*}
\alpha \triangleright_{s e} \gamma \stackrel{\text { def }}{\Leftrightarrow} \gamma \vdash_{\Sigma} \alpha \& \min (\gamma) \subseteq \min (\alpha) \tag{6}
\end{equation*}
$$

In words, an explanation of $\alpha$ is a preferred one if all its minimal models are also minimal for $\alpha$. Notice that $\alpha \triangleright_{s e} \gamma$ if and only if $C(\alpha) \subseteq C(\gamma)$ and $\gamma \vdash_{\Sigma} \alpha$. More details and motivations about this notion are given in $[26, \S 4]$ (see also the explanatory consequence relations defined in [11, 12]).

Our last variation of (5) is the following.

$$
\begin{equation*}
\alpha \triangleright_{n c} \gamma \stackrel{\text { def }}{\Leftrightarrow} \gamma \vdash_{\Sigma} \alpha \& \bmod (\gamma) \cap \min (\alpha) \neq \emptyset \tag{7}
\end{equation*}
$$

for any pair of consistent (with $\Sigma$ ) formulas $\alpha$ and $\gamma$. This condition is equivalent to $\gamma \vdash_{\Sigma} \alpha$ and $\alpha \nprec \neg \gamma$, so we call these explanations nonmonotonically consistent. This relation is somewhat similar to the weak confirmatory consequence relations defined in [11, 12].

The five postulates stated in the introduction are not enough to distinguish all the examples we have presented so far. For instance, $\triangleright_{\text {se }}$ and $\triangleright_{\Pi}$ both satisfy (a), (b), (d) and (e) but not (c).

### 2.4 Mutually incomparable worlds

The ideas used in $\S 2.3$ suggests a different way of selecting explanations. As before, let $\sqsubset$ be a strict partial order over $\bmod (\Sigma)$. A subset $A$ of $\bmod (\Sigma)$ is called a $\sqsubset$-antichain (or just antichain when no confusion about which $\sqsubset$ is being used) if the models in $A$ are mutually $\sqsubset$-incomparable. For example, $\min (\alpha)$ is an antichain for any formula $\alpha$. When the incomparability of two worlds is interpreted as some sort of "independency", then the maximal antichains provide a natural tool to select explanations. In fact, by identifying formulas with their set of models, let $\prec_{m a}$ be the reversed strict inclusion $\supset$ between antichains. Let $\triangleright_{m a}$ be the explanatory relation given by (1) with $\prec=\prec_{m a}$. Then we have

$$
\begin{equation*}
\alpha \triangleright_{m a} \gamma \stackrel{\text { def }}{\Leftrightarrow} \bmod (\Sigma \cup\{\gamma\}) \text { is a maximal antichain of } \bmod (\Sigma \cup\{\alpha\}) \tag{8}
\end{equation*}
$$

A variation of (8) is as follows

$$
\begin{equation*}
\alpha \triangleright_{\operatorname{mac}} \gamma \stackrel{\text { def }}{\Leftrightarrow} \bmod (\Sigma \cup\{\gamma\}) \text { is an antichain of } \bmod (\Sigma \cup\{\alpha\}) \text { of maximal cardinality } \tag{9}
\end{equation*}
$$

As before, we define $\prec_{\text {mac }}$ on antichains by $\delta \prec_{\text {mac }} \gamma$ when the cardinality of $\delta$ is strictly larger than the cardinality of $\gamma$. Then $\triangleright_{\text {mac }}$ is the explanatory relation associated to $\prec_{\text {mac }}$.

These two relations cannot be distinguished by the five postulates stated in the introduction, since both satisfy (a), (b) and (e) but neither (c) nor (d).

## 3 Structural rules for explanatory relations

Now we will recall the structural rules for explanatory reasoning introduced in [26].

| E-CM | If $\alpha \triangleright \gamma$ and $\gamma \vdash_{\Sigma} \beta$, then $(\alpha \wedge \beta) \triangleright \gamma$ |
| :--- | :--- |
| E-C-Cut | If $(\alpha \wedge \beta) \triangleright \gamma$ and for all $\delta\left[\alpha \triangleright \delta \Rightarrow \delta \vdash_{\Sigma} \beta\right]$, then $\alpha \triangleright \gamma$ |
| E-R-Cut | If $(\alpha \wedge \beta) \triangleright \gamma$ and there is $\delta$ such that $\delta \vdash_{\Sigma} \beta$ and $\alpha \triangleright \delta$, then $\alpha \triangleright \gamma$ |
| LOR | If $\alpha \triangleright \gamma$ and $\beta \triangleright \gamma$, then $(\alpha \vee \beta) \triangleright \gamma$ |
| E-DR | If $\alpha \triangleright \gamma$ and $\beta \triangleright \delta$, then $(\alpha \vee \beta) \triangleright \gamma$ or $(\alpha \vee \beta) \triangleright \delta$ |
| ROR | If $\alpha \triangleright \gamma$ and $\alpha \triangleright \delta$, then $\alpha \triangleright(\gamma \vee \delta)$ |
| RS | If $\alpha \triangleright \gamma, \gamma^{\prime} \vdash_{\Sigma} \gamma$ and $\gamma^{\prime} \not{ }_{\Sigma} \perp$, then $\alpha \triangleright \gamma^{\prime}$ |

Some justifications and predecessors of these rules were given in [26] (see also [11, 12]). Nevertheless, for the sake of completeness, we will make some brief comments about the intuition behind them. ${ }^{5}$

LOR and ROR are, respectively, the standard rules for the introduction of $\vee$ in the left and in the right of a relation. RS and E-CM are, respectively, strengthening rules for the right and the left. We call E-CM Explanatory Cautious Monotony since it reminds a similar rule for consequence relations. The rule E-DR, Explanatory Disjunctive Rationality, which is mainly technical, is clearly stronger than LOR. When the selection of explanations is based on a sort of syntactic simplicity criterion, it is unreasonable to have RS (a typical example is $\triangleright_{l i t}$ ).

Explanatory Cautious Cut (E-C-Cut) and Explanatory Rational Cut (E-R-Cut) are cut rules for observations. They play an important role in our setting. As we will show, they are tightly related to a selection mechanism. In fact, a failure of full explanatory cut (i.e. $\alpha \wedge \beta \triangleright \gamma$ but $\alpha \not$ d $_{\text {) says }}$ that there is some part $\beta$ of the observation $\alpha \wedge \beta$ which is so relevant for explaining the whole observation that it cannot be ignored. This difference, between the sets of preferred explanations of the whole observation and part of it, reflects an implicit preference criterion. These rules have not been considered, as far as we know, in other studies of explanatory reasoning. They are weak forms of the Converse entailment rules considered by Flach [11, 12] which are, essentially, equivalent to our full explanatory cut rule.

Example 3.1. Consider the relations $\triangleright_{\Pi}, \triangleright_{p}$ and $\triangleright_{\text {lit }}$ defined in §2.1 and §2.2. It is easy to verify that $\mathrm{E}-\mathrm{CM}$ and $\mathrm{E}-\mathrm{C}-\mathrm{Cut}$ hold for these three explanatory relations. We have already seen that LOR holds for $\triangleright_{\Pi}$ but it fails for the other two relations. On the other hand, it is not difficult to verify that $\triangleright_{p}$ satisfies $\mathrm{E}-\mathrm{R}$-Cut (essentially because $\prec_{p}$ is total) but we will see below that $\triangleright_{l i t}$ does not. This shows a basic structural difference between them. None of these relations satisfy RS.

[^3]To see that $\triangleright_{\text {lit }}$ does not satisfy E-R-Cut, let $\alpha=a \vee b, \beta=a \vee c, \gamma=c \wedge b$ and $\delta=a$. Then $(\alpha \wedge \beta) \triangleright_{l i t}(c \wedge b), \alpha \triangleright_{l i t} a, a \vdash \beta$. But $\alpha \downarrow_{l i t}(c \wedge b)$ precisely because there is another explanation of $\alpha$, namely $b$, simpler than $c \wedge b$.

The rule E-DR is not easy to satisfy since, as we will see, it imposes a strong constraint on the preference relation. For instance, for the relation $\triangleright_{\Pi}$ take $\Sigma$ to be $a \rightarrow b$ and $\Pi$ to be $\Sigma$ together with $b \rightarrow a$. Then E-DR fails. In fact, from (3) it follows that $b \triangleright_{\Pi} a, c \triangleright_{\Pi} c,(b \vee c) \not \wp_{\Pi} a$ and $(b \vee c) \not{ }_{\Pi} c$. Finally, it also follows easily that $\triangleright_{\Pi}$ satisfies E-R-Cut ${ }^{6}$

Table 1 summarizes the postulates satisfied by the examples of explanatory relations given in $\S 2$. The symbols $\sqrt{ }(f)$ and $\sqrt{ }(m)$, respectively, means that the postulate is satisfied when the partial order $\sqsubset$ over $\bmod (\Sigma)$ is filtered (see definition $\left.4.4^{7}\right)$ and modular (i.e. $\sqsubset$ is a total pre-order).

|  | E-CM | E-C-Cut | LOR | E-R-Cut | ROR | E-DR | RS |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\triangleright_{c}$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }(\mathrm{~m})$ | $\sqrt{ }$ | $\sqrt{ }(\mathrm{f})$ | $\sqrt{ }$ |
| $\triangleright_{s e}$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }(\mathrm{~m})$ | $\sqrt{ }$ | $\sqrt{ }(\mathrm{f})$ | $\times$ |
| $\triangleright_{n c}$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }(\mathrm{f})$ | $\sqrt{ }(\mathrm{m})$ | $\sqrt{ }$ | $\sqrt{ }(\mathrm{f})$ | $\times$ |
| $\triangleright_{\Pi}$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }$ | $\times$ | $\times$ |
| $\triangleright_{m a c}$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }$ | $\times$ | $\times$ | $\times$ |
| $\triangleright_{m a}$ | $\sqrt{ }$ | $\sqrt{ }$ | $\sqrt{ }$ | $\times$ | $\times$ | $\times$ | $\times$ |
| $\triangleright_{p}$ | $\sqrt{ }$ | $\sqrt{ }$ | $\times$ | $\sqrt{ }$ | $\times$ | $\times$ | $\times$ |
| $\triangleright_{l i t}$ | $\sqrt{ }$ | $\sqrt{ }$ | $\times$ | $\times$ | $\times$ | $\times$ | $\times$ |

Table 1: Postulates satisfied by some examples of explanatory relations

It is clear that two logically equivalent (modulo $\Sigma$ ) observations have the same set of (possible) explanations Expla $(\cdot)$. Therefore one would expect that they will also have the same preferred explanations. It is also clear, that in order to have non-trivial relations it is necessary to guarantee the existence of at least one preferred explanation for each $\alpha$ consistent with $\Sigma$. Finally, in (1) if a formula is a $\prec$-minimal explanation for some observation $\alpha$, then it is necessarily minimal for itself. These considerations are the content of the following postulates. $\mathrm{LLE}_{\Sigma}$ and $\mathrm{E}-\mathrm{Con}_{\Sigma}$ stand respectively for Left Logical Equivalence and Explanatory Consistency Preservation.

| $\operatorname{LLE}_{\Sigma}$ | $\vdash_{\Sigma} \alpha \leftrightarrow \alpha^{\prime}$ and $\alpha \triangleright \gamma$, then $\alpha^{\prime} \triangleright \gamma$ |
| :--- | :--- |
| E-Con $_{\Sigma}$ | if $\vdash_{\Sigma} \neg \alpha$, then there is $\gamma$ such that $\alpha \triangleright \gamma$ |
| E-Reflexivity | If $\alpha \triangleright \gamma$, then $\gamma \triangleright \gamma$ |

In [26] the following classification of explanatory relations based on these postulates was introduced.

[^4]Definition 3.2. Let $\Sigma$ be a background theory and $\triangleright$ be an explanatory relation. We say that $\triangleright$ is E-cumulative if it satisfies $\mathrm{LLE}_{\Sigma}, \mathrm{E}-\mathrm{CM}, \mathrm{E}-\mathrm{C}-\mathrm{Cut}$ and $\mathrm{E}-\mathrm{Con}_{\Sigma}{ }^{8}$. We say that $\triangleright$ is E-preferential if it is E-cumulative and also satisfies RS. $\triangleright$ is $\mathbf{E}$-disjunctive rational if it is E-preferential and in addition satisfies E-DR. $\triangleright$ is E-rational if it is E-preferential and in addition satisfies E-R-Cut.

We briefly recall the main intuition behind this classification (see more details in [26]). To each explanatory relation $\triangleright$ we associate a consequence relation $\vdash_{a b}$ as follows:

$$
\begin{equation*}
\alpha \vdash_{a b} \beta \text { if } \Sigma \cup\{\gamma\} \vdash \beta \text { for each } \gamma \text { such that } \alpha \triangleright \gamma \tag{10}
\end{equation*}
$$

We read $\alpha \sim_{a b} \beta$ as "when $\alpha$ is observed, then normally $\beta$ is also present". Let $C_{a b}(\alpha)=\{\beta$ : $\left.\alpha \vdash_{a b} \beta\right\}$. Formulas in $C_{a b}(\alpha)$ are called the abductive consequences of $\alpha$.

The explanatory postulates presented above were introduced in order to assure that $\sim_{a b}$ satisfies the well known postulates for consequence relations studied by Kraus, Lehmann and Magidor $[18,19]$ and many others [13, 15]. In fact, the relation $\sim_{a b}$ is preferential (respectively disjunctive rational and rational) when $\triangleright$ is E-preferential (respectively E-disjunctive rational and E-rational) [26]. All explanatory relations found in this paper will be at least E-cumulative ${ }^{9}$.

The analysis of $\psi_{a b}$ made evident that there is a distinguished family of explanatory relations, called causal, for which $\vdash_{a b}$ completely captures the explanatory relation, in the sense that $\triangleright$ can recovered from $\sim_{a b}[26, \S 3.1]$. More precisely, an explanatory relation is causal if the following holds

$$
\begin{equation*}
\alpha \triangleright \gamma \text { iff } C_{a b}(\alpha) \subseteq C n(\Sigma \cup \gamma) \tag{11}
\end{equation*}
$$

Notice that the only if part in (11) always holds. The interpretation of this condition is as follows: The preferred explanations of an observation $\alpha$ are exactly those explanations which account also for other facts that are normally present when $\alpha$ is observed ${ }^{10}$.

It is easy to verify that any causal relation satisfies RS and ROR. Moreover, in a finite language, an explanatory relation is causal if and only if these two postulates are satisfied [26]. The reader can take for the rest of the paper RS and ROR as the defining conditions of a causal relation. The relation $\triangleright_{c}$ defined in $\S 2.3$ is the typical causal explanatory relation. In fact, any causal explanatory relation is essentially of that form and thus they have a semantics in terms of minimal models [26, §3.1].

We end this section with a technical notion which is helpful for dealing with infinite languages. Some of our results are proved for explanatory relations that behave as if the language is finite. We could have restricted ourselves to work only with finite languages, but we have introduced these new postulates, of course, to gain in generality but also because these postulates seem to be very appropriate for explanatory reasoning in situations where there are infinite many possible observations and explanations. These postulates essentially say that every observation has at most finitely many preferred explanations and every formula is a preferred explanation of at most finitely many observations.

[^5]Definition 3.3. $A$ set of formulas $A$ is said to have a lower bound ${ }^{11}$ (in $A$ with respect to $\Sigma$ ) if there are finitely many formulas $\alpha_{1}, \ldots, \alpha_{n} \in A$ such that for all $\alpha \in A, \alpha \vdash_{\Sigma}\left(\alpha_{1} \vee \cdots \vee \alpha_{n}\right)$.

Definition 3.4. An explanatory relation $\triangleright$ is said to be logically finite on the right (RLF) if for every formula $\alpha$ the set $\{\gamma: \alpha \triangleright \gamma\}$ is either empty or it has a lower bound. $\triangleright$ is said to be logically finite on the left (LLF) if for every formula $\gamma$ the set $\{\alpha: \alpha \triangleright \gamma\}$ is either empty or it has a lower bound. $\triangleright$ is said to be logically finite if it satisfies RLF or LLF.

## 4 Structural properties for preference relations

In this section we introduce some notions concerning preference criteria. In section 5 we will show that these notions are the formal counterpart of the structural rules for explanatory relations introduced in the previous section.

Let $\prec$ be an irreflexive binary relation over a set $S$ and $A \subseteq S$. An element $a \in A$ is a $\prec$-minimal element of $A$ if there is no $b \in A$ with $b \prec a$. The minimal elements of a set $A$ will be denoted by $\min (A, \prec)$ and when there is no confusion about which preference relation $\prec$ is being used, we will just write $\min (A)$.

Definition 4.1. A preference relation $\prec$ will be any binary irreflexive relation $\prec$ over the collection of formulas ${ }^{12}$.

Of course, some extra requirements must be imposed on a preference relation. For instance, for a given a preference relation $\prec$, the explanatory relation defined by $(1)$ is meaningful if $\min (\operatorname{Expla}(\alpha), \prec$ ) is not empty. For a transitive and irreflexive relation $\prec$ on a finite set $S$, it is clear that $\min (A)$ is not empty for any non-empty $A \subseteq S$. However, there are non-transitive relations satisfying this. As we will see, the transitivity of the preference relation is not needed in most of our results. Transitivity does not seem to be an essential part of the selection process but it is clearly an extra feature that makes things easier. We will use a property, called smoothness, which locally guarantees the existence of minimal elements. This notion has been already used in the study of consequence relations [18].

Definition 4.2. Let $\prec$ be a irreflexive binary relation over a set $S$. We say that a subset $A \subseteq S$ is smooth if for every $a \in A$ either $a$ is minimal in $A$ or there is $b \in A$ with $b \prec a$ and $b$ minimal in A. A preference relation $\prec$ is called smooth, if Expla $(\alpha)$ is smooth for every formula $\alpha$.

The following observation clarifies the meaning of smoothness. Let $A \subseteq B \subseteq S$, then clearly $\min (B) \cap A \subseteq \min (A)$. Suppose now that $\min (B) \subseteq A$, hence $\min (B) \subseteq \min (A)$. It is reasonable then to expect that $\min (A)=\min (B)$. This is true when $B$ is smooth since, in this case, $\min (A) \subseteq$ $\min (B)$.

There are several stronger forms of smoothness that will be useful: disjunctively smooth, filtered and modular (definitions are given below). Roughly speaking, these four notions are related to the left hand side of an explanatory relation. Namely E-C-Cut, LOR, E-DR and E-R-Cut, respectively.

[^6]For the right hand side of an explanatory relation we have considered only two postulates: RS and ROR. The corresponding properties for preference relations are denoted by $E_{1}$ and $E_{2}$ and will be introduced at the end of this section.

If $\prec$ is a preference relation such that $\nabla_{\prec}$ satisfies LOR, then it is easy to check that the following holds for all $\alpha$ and $\beta$

$$
\begin{equation*}
\min (\operatorname{Expla}(\alpha), \prec) \cap \min (\operatorname{Expla}(\beta), \prec) \subseteq \min (\operatorname{Expla}(\alpha \vee \beta), \prec) \tag{12}
\end{equation*}
$$

This property can be stated in a way which also generalizes smoothness as follows.
Definition 4.3. A preference relation $\prec$ is called disjunctively smooth if for every $\alpha, \beta$ and $\gamma \vdash_{\Sigma} \alpha \vee \beta$ such that $\gamma \notin \min (\operatorname{Expla}(\alpha \vee \beta), \prec)$, there is $\delta \prec \gamma$ such that $\delta \in \min (\operatorname{Expla}(\alpha), \prec)$ or $\delta \in \min (\operatorname{Expla}(\beta), \prec)$.

The next notion is motivated by the notion of a filtered preferential model introduced in the study of consequence relations [13].

Definition 4.4. Let $\prec$ be a relation over a set $S$. $A$ subset $A \subseteq S$ is said to be $\prec$-filtered if for every $a, b \in A$ such that $a, b \notin \min (A, \prec)$, there is $c \in \min (A, \prec)$, such that $c \prec a$ and $c \prec b$. An order $\sqsubset \operatorname{over} \bmod (\Sigma)$ is said to be filtered, if $\bmod (\Sigma \cup\{\alpha\})$ is $\sqsubset$-filtered for every formula $\alpha$. $A$ preference relation $\prec$ is called filtered, if Expla $(\alpha)$ is $\prec$-filtered, for every formula $\alpha$

For instance, the relation $\prec_{\Pi}$ is filtered. In fact, given $\gamma, \delta \in \operatorname{Expla}(\alpha)$ but not minimal, then $\alpha \prec_{\Pi} \gamma$ and $\alpha \prec_{\Pi} \delta$. Any filtered preference relation is obviously smooth. The properties of being filtered and disjunctively smooth are independent notions as the following example shows.

Example 4.5. Consider the relations $\prec_{p}$ and $\prec_{m a}$ defined in §2. We will show that $\prec_{p}$ is filtered but not disjunctively smooth and that the opposite happens for $\prec_{m a}$.

We have already shown that $\triangleright_{p}$ does not satisfies LOR which is the same to saying that $\prec_{p}$ is not disjunctively smooth. To see that $\prec_{p}$ is filtered, suppose $\gamma, \delta$ are formulas in Expla( $\alpha$ ) but are not minimal. Let $\rho$ be $a \prec_{p}$-minimal element of Expla( $\alpha$ ), then necessarily $|\rho|<|\gamma|$ and $|\rho|<|\delta|$.

Now we will analyze $\prec_{m a}$. Since $\triangleright_{m a}$ satisfies LOR, then $\prec_{m a}$ is disjunctively smooth. Consider the following situation. Let $\Sigma$ be such that its models are exactly $N, M, P$ and $Q$. Let $\sqsubset$ be given by the following

$$
P \sqsubset N, \quad P \sqsubset M, \quad Q \sqsubset N, \quad Q \sqsubset M
$$

The antichains are $\{N\},\{M\},\{P\},\{Q\},\{N, M\}$ and $\{P, Q\}$. So the order $\prec_{m a}$ is given by

$$
\{N, M\} \prec_{m a}\{N\}, \quad\{N, M\} \prec_{m a}\{M\}, \quad\{P, Q\} \prec_{m a}\{P\}, \quad\{P, Q\} \prec_{m a}\{Q\}
$$

Let $\alpha$ be a formula whose models are $N, M, P$ and $Q$. The $\prec_{\text {ma-minimal explanations of } \alpha \text { are }}$ $\{N, M\}$ and $\{P, Q\}$. To see $\prec_{m a}$ is not filtered, let $\gamma$ and $\delta$ be, respectively, $\{N\}$ and $\{P\}$ and let $\alpha$ be as above.

For a filtered $\sqsubset$ the explanatory relations $\triangleright_{c}, \triangleright_{n c}$ and $\triangleright_{s e}($ defined in $\S 2.3)$ satisfied E-DR. The analogous result for consequence relations says that a filtered preferential model induces a consequence relation satisfying Disjunctive Rationality ${ }^{13}$. However, a filtered preference relation $\prec$ does not necessarily induce an explanatory relation satisfying E-DR. A strengthening of the property of being filtered is needed which looks very technical.

[^7]Definition 4.6. A preference relation $\prec$ is said to be strongly filtered if for every $\alpha$ and every $\gamma, \delta \in \operatorname{Expla}(\alpha)$ such that $\gamma, \delta \notin \min (\operatorname{Expla}(\alpha), \prec)$, there is $\rho \in \min (\operatorname{Expla}(\alpha), \prec)$ such that $\rho \prec \gamma$, $\rho \prec \delta$ and moreover $\rho^{\prime} \prec \gamma, \rho^{\prime} \prec \delta$ for every $\rho^{\prime} \vdash_{\Sigma} \rho$ consistent with $\Sigma$.
Remark 4.7. If $\prec$ is a transitive and strongly filtered relation, then $\prec$ is disjunctively smooth. In fact, suppose $\gamma \vdash_{\Sigma} \alpha \vee \beta$ and $\gamma \notin \min ($ Expla $(\alpha \vee \beta)$, $\prec)$. Let $\rho$ be such that $\rho \prec \gamma$ as in the definition of a strongly filtered relation. Consider the following cases. (i) $\rho \vdash_{\Sigma} \alpha$ or $\rho \vdash_{\Sigma} \beta$. It follows from the transitivity of $\prec$ that $\rho \in \min (\operatorname{Expla}(\alpha), \prec)$ or $\rho \in \min (\operatorname{Expla}(\beta), \prec)$. (ii) Suppose $\rho \Vdash_{\Sigma} \alpha$ and $\rho \nvdash_{\Sigma} \beta$. Then $\rho \wedge \alpha$ is consistent with $\Sigma, \rho \wedge \alpha \prec \gamma\left(\right.$ as $\left.\rho \wedge \alpha \vdash_{\Sigma} \rho\right)$ and $\rho \wedge \alpha \in \min (\operatorname{Expla}(\alpha), \prec)$ (by transitivity).

This version of the property of being filtered imposes strong constraints on the preference relation. For instance, the relation $\prec_{\Pi}$ is strongly filtered if and only if $\Pi$ is complete (i.e. $\Pi \vdash \alpha$ or $\Pi \vdash \neg \alpha$ for every formula $\alpha$ ).

The notions introduced so far are satisfied by partial orders. The next one, called modularity, imposes linearity. It has been used in the study of inference relations [19].
Definition 4.8. A relation $\prec$ on a set $E$ is said to be modular if there exists a linear order $<$ on some set $\Omega$ and a function $r: E \rightarrow \Omega$ such that $a \prec b$ iff $r(a)<r(b)$.

Some examples of modular relations are $\prec_{p}$ and $\prec_{\text {mac }}$. For transitive relations, modularity is equivalent to the following property. Let $a, b, c$ in $E$ such that $a$ and $b$ are $\prec$-incomparable and $a \prec c$, then $b \prec c$. Modularity does not imply smoothness. But a smooth and modular preference relation is easily seen to be filtered, but not necessarily disjunctively smooth (an example is $\prec_{p}$ which is not disjunctively smooth because its associate explanatory relation $\triangleright_{p}$ does not satisfy LOR).

Finally we introduce the properties that correspond to RS and ROR.
$\mathrm{E}_{1}$ If $\gamma \nvdash_{\Sigma} \perp, \gamma \vdash \gamma^{\prime}$ and $\delta \prec \gamma$, then $\delta \prec \gamma^{\prime}$.
$\mathrm{E}_{2}$ If $\rho \prec \gamma \vee \delta$, then $\rho \prec \gamma$ or $\rho \prec \delta$.
These two properties are stated without mentioning observations. In a sense, they are related to the lattice of formulas. They are somewhat similar to the properties studied by Freund [13].

None of the examples of preference relations defined in $\S 2$ satisfied $E_{1}$. An example of one which does satisfy it will be given in $\S 6.1$. The order $\prec_{\Pi}$ satisfies $E_{2}$ and $\prec_{l i t}$ does not.

It is not difficult to verify that a smooth modular relation satisfying $\mathrm{E}_{1}$ is strongly filtered. There are examples of filtered relations satisfying $E_{1}$ which are not strongly filtered (see $\S 6.1$ ).

Figure 1 summarizes the interconnections between all notions introduced in this section. An arrow means logical implication and in the case that an extra property is needed for the validity of the implication we indicate the extra property as a label above the corresponding arrow.

## 5 Representation theorems

In this section we will present the main results. We recall that to each preference relation $\prec$ is associated an explanatory relation as follows

$$
\begin{equation*}
\alpha \triangleright_{\prec} \gamma \stackrel{\text { def }}{\Leftrightarrow} \gamma \in \min (\operatorname{Expla}(\alpha), \prec) \tag{13}
\end{equation*}
$$



Figure 1: Properties for preference relations

It should be clear by now that the issues we want to address are the following. First, we will analyze which structural properties of $\prec$ (like being smooth, filtered, modular, etc.) are sufficient to have that the associated relation $\nabla_{\prec}$ satisfies a given postulate. Second, given an explanatory relation $\triangleright$, we would like to know when there is a preference relation $\prec$ such that $\triangleright=\nabla_{\prec}$. In this case, we will say that $\triangleright$ is represented by $\prec$ or simply that it is representable. A fairly detailed picture will emerge from the answers to these questions.

The main results are the following four theorems. In the appendix $A$ the reader will find the proofs and some technical results (which are needed in the proofs of the main theorems) showing some finer aspects of the relationship between postulates and structural properties of preference relations.

Theorem 5.1. Let $\triangleright$ be a logically finite explanatory relation. The following are equivalent.
(i) $\triangleright$ is E-cumulative and satisfies LOR.
(ii) $\triangleright$ is representable by a disjunctively smooth preference relation.

Theorem 5.2. Assume the language is finite and let $\triangleright$ be an explanatory relation. The following are equivalent
(i) $\triangleright$ is E-disjunctive rational (i.e E-cumulative $+\mathrm{RS}+\mathrm{E}-\mathrm{DR}$ ).
(ii) $\triangleright$ is representable by a strongly filtered, transitive preference relation that satisfies $\mathrm{E}_{1}$.

It turns out that when E-R-Cut also holds, it is unnecessary to assume logical finiteness (we recall that E-cumulative $+\mathrm{RS}+$ E-R-Cut implies E-DR [26, Proposition 2.14]). More precisely we have the following

Theorem 5.3. Let $\triangleright$ be an explanatory relation. The following are equivalent
(i) $\triangleright$ is E-rational (i.e E-cumulative $+\mathrm{RS}+\mathrm{E}-\mathrm{R}-\mathrm{Cut})$.
(ii) $\triangleright$ is representable by a smooth and modular preference relation that satisfies $\mathrm{E}_{1}$.

As we have mentioned in $\S 3$, causal explanatory relations (i.e those satisfying RS and ROR) are a distinguished collection of explanatory relations. The properties of the associate consequence relation $\sim_{a b}$ are crucial to study this type of explanatory relations. The next results is about causal
relations. Its proof uses in a crucial way the properties of $\sim_{a b}$, since it is based on the representation of $\sim_{a b}$ in terms of preferential models.

Theorem 5.4. Suppose the language is finite and let $\triangleright$ be a E-cumulative explanatory relation. (i) If $\triangleright$ satisfies RS and ROR, then it is representable by a smooth preference relation.
(ii) $\triangleright$ satisfies $\mathrm{RS}, \mathrm{ROR}$ and LOR iff it is representable by a disjunctively smooth preference relation satisfying $\mathrm{E}_{1}$ and $\mathrm{E}_{2}$.

The harder part of the proof of such results is to show the representability of an explanatory relation. The key fact for doing this will be to associate to each explanatory relation a preference relation, called the essential relation. To define it we need an auxiliary notion.

Definition 5.5. Let $\triangleright$ be an explanatory relation. We will say that a formula $\gamma$ is admissible for $\triangleright$ (or just admissible) if $\alpha \triangleright \gamma$ for some formula $\alpha$.

Admissible formulas are consistent with $\Sigma$. They play in our paper the same role as normal models in $[18,22,27]$. If $\triangleright$ satisfies E-CM and E-C-Cut, then it is easy to check that E-Reflexivity holds, that is to say, $\gamma \triangleright \gamma$ for every admissible $\gamma$.

The following definition was motivated by the results in [27].
Definition 5.6. Let $\triangleright$ be an explanatory relation. The essential preference relation associated to $\triangleright$ is denoted by $\prec^{e}$ and defined as follows. Let $\gamma$ be an admissible formula

$$
\gamma \prec^{e} \delta \stackrel{\text { def }}{\Leftrightarrow} \forall \alpha\left(\alpha \triangleright \delta \Rightarrow \gamma \vdash_{\Sigma} \alpha\right) .
$$

Remark 5.7. A more precise notation would have been something like $\prec_{\triangleright}^{e}$ so that to stress the dependency of $\prec^{e}$ with the explanatory relation $\triangleright$. If it is necessary, to avoid any confusion, we will add some extra subscript.

Admissible formulas are the only relevant formulas for the definition of $\prec^{e}$, since $\gamma \prec^{e} \delta$ for every admissible $\gamma$ and non-admissible $\delta$.

The main part of the proofs will consist in showing that the following holds (under the hypothesis in part (i) of each of the theorems above)

$$
\begin{equation*}
\alpha \triangleright \gamma \quad \text { iff } \quad \gamma \in \min \left(\operatorname{Expla}(\alpha), \prec^{e}\right) \tag{14}
\end{equation*}
$$

In other words, $\prec^{e}$ represents $\triangleright$.
The "only if" direction in (14) is always true. Moreover, we have the following straightforward observation saying that if an explanatory relation is representable by a preference relation, then $\prec^{e}$ is the largest possible one.

Proposition 5.8. Suppose $\prec$ is a preference relation and $\gamma$ is an admissible formula for $\triangleright_{\prec}$. If $\gamma \prec \delta$, then $\gamma \prec^{e} \delta$. Moreover, if $\prec$ is smooth, then $\prec^{e}$ also represents $\triangleright_{\prec}$.

In the next section we will show examples of explanatory relations admitting at least two different representations. On the other hand, there are also E-cumulative explanatory relations which are not representable by its associated preference relation (an example is given in the appendix, see A.19).

The next result shows that $\prec^{e}$ preserves a bit of the lattice structure on formulas. Its proof is straightforward.

Proposition 5.9. Let $\triangleright$ be an explanatory relation. Then the following holds:
(i) Let $\gamma, \gamma^{\prime}$ and $\delta$ be admissible formulas such that $\gamma \vdash \gamma^{\prime}$. If $\gamma \prec^{e} \delta$, then $\gamma^{\prime} \prec^{e} \delta$.
(ii) Let $\gamma$ and $\delta$ be admissible formulas. If $(\delta \vee \gamma) \prec^{e} \gamma$, then $\delta \prec^{e} \gamma$.

The next result completes the proof of the main theorems.
Theorem 5.10. Let $\prec$ be a binary irreflexive relation. Then $\nabla_{\prec}$ satisfies E-Reflexivity and E-CM. If in addition $\prec$ is smooth, then $\nabla_{\prec}$ is E-cumulative.

Since any partial order over a finite set is smooth, we immediately get the following
Corollary 5.11. Suppose the language is finite and $\prec$ is a strict partial order over formulas. Then $D_{\prec}$ is E-cumulative.

We finish this section showing that some special explanatory relations are representable assuming neither LOR nor logical finiteness.

An explanatory relation is said to be fully reflexive if $\alpha \triangleright \alpha$ for every formula $\alpha$ consistent with $\Sigma$. This might not seem a natural condition, since $\triangleright$ is intended to capture the notion of a preferred explanation. Nevertheless, there are some interesting examples of explanatory relations of this type (for instance $\triangleright_{\Pi}, \triangleright_{n c}$ and $\triangleright_{s e}$ ). It is easy to verify that full reflexivity together with E-CM implies E-C-Cut.
Proposition 5.12. Let $\triangleright$ be an explanatory relation satisfying $\mathrm{LLE}_{\Sigma}, \mathrm{E}-\mathrm{CM}$ and full reflexivity. Then $\prec^{e}$ is a filtered preference relation representing $\triangleright$.

Our next proposition is more interesting, it shows that E-R-Cut suffices to get the representability of an explanatory relation.
Proposition 5.13. Let $\triangleright$ be an E-cumulative explanatory relation satisfying E-R-Cut. Then $\prec^{e}$ is smooth, filtered and represents $\triangleright$.

Our last observation says that full cut is, as expected, a very strong requirement since it trivializes the preference relation. In other words, instances where full cut fails reflect the implicit preference criterion. Its proof is straightforward.

Proposition 5.14. Suppose $\triangleright$ is an explanatory relation satisfying full cut (i.e. if $(\alpha \wedge \beta) \triangleright \gamma$, then $\alpha \triangleright \gamma$ ). Then $\gamma \prec^{e} \delta$ if and only if $\gamma$ is admissible and $\delta$ is not admissible.

## 6 The examples revisited

In this section we will analyze the essential relation associated with the explanatory relations given in $\S 2$. The first three examples are the most interesting since these explanatory relations were not defined in terms of a preference relation over formulas. However, as we will show, they are indeed represented by their associate essential relations. Also we will show that Freund's preferential orders [13] correspond to the essential relation of a particular type of explanatory relations. This includes also the notions of expectation orders [15] and possibilistic orders [7] since they are preferential orders.

The other two examples correspond to explanatory relations already defined by preference relations, so the fact that they are also represented by their associated essential relations gives apparently not much additional information about them (see proposition 5.8). Nevertheless, they will illustrate some results used in the proof of the main theorems (see A. 3 in the appendix).

### 6.1 Causal explanations

Recall that for a given partial order over $\bmod (\Sigma)$ we define

$$
\begin{equation*}
\alpha \triangleright_{c} \gamma \stackrel{\text { def }}{\Leftrightarrow} \bmod (\Sigma \cup\{\gamma\}) \subseteq \min (\alpha) \tag{15}
\end{equation*}
$$

for any pair of consistent (with $\Sigma$ ) formulas $\alpha$ and $\gamma$. This method always yields E-cumulative explanatory relations that moreover satisfies ROR, RS and LOR. By theorem 5.4 (or also theorem 5.1) $\triangleright_{c}$ is represented by its associated essential relation, denoted by $\prec_{c}^{e}$. The next characterization of $\prec_{c}^{e}$ in terms of $\sqsubset$ is given in the appendix (see A.16). For $\gamma$ and $\delta$ admissible

$$
\gamma \prec_{c}^{e} \delta \text { iff } \exists N \models \delta \exists M \models \gamma \text { such that } M \sqsubset N
$$

The representation theorem 5.1 guarantees that $\prec_{c}^{e}$ is disjunctively smooth and moreover it is also filtered and satisfies $\mathrm{E}_{1}$ and $\mathrm{E}_{2}$ (by A. 6 and A.17).

In general, the relation $\prec_{c}^{e}$ is not transitive. In fact, it is easy to find an example of a partial order $\sqsubset$ and formulas $\gamma$ and $\delta$ such that $\gamma \prec_{c}^{e} \delta$ and also $\delta \prec_{c}^{e} \gamma$.

If $\sqsubset$ is filtered (see definition 4.4), then $\triangleright_{c}$ satisfies E-DR and therefore in this case $\prec_{c}^{e}$ is transitive and strongly filtered (by A.4). Finally, if $\sqsubset$ is modular (i.e. a total pre-order), then $\triangleright_{c}$ is E-rational and therefore $\prec_{c}^{e}$ is also modular (by A.9).

### 6.2 Strong epistemic explanations

Recall the definition of $\triangleright_{s e}$

$$
\begin{equation*}
\alpha \triangleright_{s e} \gamma \stackrel{\text { def }}{\Leftrightarrow} \gamma \vdash_{\Sigma} \alpha \& \min (\gamma) \subseteq \min (\alpha) \tag{16}
\end{equation*}
$$

for any pair of consistent (with $\Sigma$ ) formulas $\alpha$ and $\gamma$. As we said, $\triangleright_{\text {se }}$ is E-cumulative and moreover satisfies ROR and LOR. However RS does not hold. Since $\triangleright_{\text {se }}$ is fully reflexive, then every formula consistent with $\Sigma$ is $\triangleright_{s e}$-admissible. The relation $\prec_{s e}^{e}$ is characterized as follows.

$$
\gamma \prec_{s e}^{e} \delta \text { iff } \exists N \in \min (\delta) \exists M \in \min (\gamma) \text { such that } M \sqsubset N
$$

So the crucial difference with $\prec_{c}^{e}$ is the notion of an admissible formula. By theorem 5.1, $\triangleright_{s e}$ is representable by $\prec_{s e}^{e}$ ( 5.12 also applies here). In general, $\prec_{s e}^{e}$ is not transitive but it is disjunctively smooth and filtered (by A.17) and it satisfies $E_{2}$ but not $E_{1}$, as RS fails.

### 6.3 Nonmonotonically consistent explanations

Recall the definition of $\triangleright_{n c}$

$$
\begin{equation*}
\alpha \triangleright_{n c} \gamma \stackrel{\text { def }}{\Leftrightarrow} \gamma \vdash_{\Sigma} \alpha \& \bmod (\gamma) \cap \min (\alpha) \neq \emptyset \tag{17}
\end{equation*}
$$

for any pair of consistent (with $\Sigma$ ) formulas $\alpha$ and $\gamma$. As we said, $\triangleright_{n c}$ is E-cumulative and moreover satisfies ROR. But as in the previous example, RS does not hold. $\triangleright_{n c}$ is also fully reflexive. It is interesting to observe that $\triangleright_{n c}$ might not satisfy LOR (and hence theorem 5.1 does not apply). However, by 5.12 it is representable by $\prec_{n c}^{e}$.

From the following characterization of $\prec_{n c}^{e}$ it is easy to check that $\prec_{n c}^{e}$ is transitive, satisfies $\mathrm{E}_{2}$ and is filtered (by A.17). We have that

$$
\begin{equation*}
\gamma \prec_{n c}^{e} \delta \text { iff } \forall N \in \min (\delta) \exists M \in \min (\gamma) \text { with } M \sqsubset N \tag{18}
\end{equation*}
$$

(To see $\Leftarrow$, let $\alpha \triangleright_{n c} \delta$ and $N \models \delta$ such that $N \in \min (\alpha)$. Let $M \models \gamma$ such that $M \sqsubset N$. Thus $M \not \models \alpha$ and so $\gamma \nvdash_{\Sigma} \alpha$. To see the converse we show its contrapositive. In fact, it suffices to verity that $\left.(\gamma \vee \delta) \triangleright_{n c} \delta\right)$ ).

Now we will show that $\prec_{n c}^{e}$ is exactly a preferential order as defined by Freund [13]. Thus when $\sqsubset$ is modular, $\prec_{n c}^{e}$ is (dually) $)^{14}$ an expectation order as defined by Gärdenfors and Makinson [15] or equivalently a possibilistic order as defined by Dubois and Prade [7].

We recall that the preferential order associated to a preferential consequence relation $\sim$ is defined by $\gamma<\delta$ if $\gamma \vee \delta \mu \neg \delta$ (Freund showed that in fact any preferential order is of this form). Now from (18) it is straightforward that $\prec_{n c}^{e}$ is the preferential order associated to the consequence relation $\mathcal{F}_{\sqsubset}$ given by (4).

### 6.4 Mutually incomparable worlds

Recall the definition of $\triangleright_{m a}$

$$
\begin{equation*}
\alpha \triangleright_{m a} \gamma \stackrel{\text { def }}{\Leftrightarrow} \bmod (\Sigma \cup\{\gamma\}) \text { is a maximal antichain of } \bmod (\Sigma \cup\{\alpha\}) \tag{19}
\end{equation*}
$$

Since $\triangleright_{m a}$ is the explanatory relation associated to a the partial order $\prec_{m a}$, then $\triangleright_{m a}$ is Ecumulative (by 5.11 ). It is easy to verify that LOR also holds, but E-DR, ROR and RS might fail (see an example below). The associated essential relation will be denoted by $\prec_{m a}^{e}$. It can be stated in terms of $\sqsubset$ as follows. Let $\gamma$ and $\delta$ be formulas whose set of models are an antichain. Then

$$
\delta \prec_{m a}^{e} \gamma \text { iff } \bmod (\Sigma \cup\{\gamma\}) \cup\{N\} \text { is an antichain for some } N \models \delta \wedge \neg \gamma
$$

(to see $\Leftarrow$, suppose $\alpha \triangleright_{m a} \gamma$ and show that $N \not \vDash \alpha$. For the other direction it is easier to show its contrapositive. In fact, just notice that $\left.(\gamma \vee \delta) \triangleright_{m a} \gamma\right)$.

Since $\triangleright_{m a}$ satisfies LOR, then $\prec_{m a}^{e}$ is disjunctively smooth but it is not filtered. The same holds for $\prec_{m a}$ as we saw in the example 4.5. That example also illustrates the difference between $\prec_{m a}$ and $\prec_{m a}^{e}$. Recall that in that example $\bmod (\Sigma)$ consists of four valuations $N, M, P$ and $Q$. They are ordered as indicated below

$$
P \sqsubset N, \quad P \sqsubset M, \quad Q \sqsubset N, \quad Q \sqsubset M
$$

The antichains are $\{N\},\{M\},\{P\},\{Q\},\{N, M\}$ and $\{P, Q\}$. The relation $\prec_{m a}^{e}$ is as follows (identifying formulas with its sets of models)

$$
\begin{array}{cccc}
\{N\} \prec_{m a}^{e}\{M\}, & \{M\} \prec_{m a}^{e}\{N\}, & \{P\} \prec_{m a}^{e}\{Q\}, & \{Q\} \prec_{m a}^{e}\{P\} \\
\{N, M\} \prec_{m a}^{e}\{N\}, & \{N, M\} \prec_{m a}^{e}\{M\}, & \{P, Q\} \prec_{m a}^{e}\{P\}, & \{P, Q\} \prec_{m a}^{e}\{Q\}
\end{array}
$$

Notice $\prec_{m a}$ consists only of the pairs in the second row above. To see that ROR fails, take $\alpha$ to be a formula whose models are $N, M, P$ and $Q$. Then $\{N, M\}$ and $\{P, Q\}$ are preferred explanations of $\alpha$ but $\{N, M, P, Q\}$ is not.

[^8]
### 6.5 Cautious and simple explanations

For $\Pi \supseteq \Sigma$ a collection of formulas we defined a preference relation by $\gamma \prec_{\Pi} \delta$ if $\delta \vdash_{\Pi} \gamma$ and $\gamma \vdash_{\Pi} \delta$. The associated explanatory relation $\triangleright_{\Pi}$ is

$$
\alpha \triangleright_{\Pi} \gamma \Leftrightarrow \gamma \nvdash_{\Sigma} \perp \& \gamma \vdash_{\Sigma} \alpha \& \alpha \vdash_{\Pi} \gamma
$$

It is easy to verify that $\triangleright_{\Pi}$ is E-cumulative, fully reflexive and also satisfies ROR, LOR and E-R-Cut. In most cases RS and E-DR fails. The associate essential relation will be denoted by $\prec_{\Pi}^{e}$. It is (in fact disjunctively) smooth and filtered (by theorem 5.13). Moreover, it is easy to check that $\prec_{\Pi}^{e}$ satisfies $E_{2}$, it is not strongly filtered and does not satisfies $E_{1}$. From the definition of the essential relation we have

$$
\delta \prec_{\Pi}^{e} \gamma \Leftrightarrow \forall \alpha\left[\gamma \vdash_{\Sigma} \alpha \& \alpha \vdash_{\Pi} \gamma \Rightarrow \delta \vdash_{\Sigma} \alpha\right]
$$

From this it follows that if $\delta \vdash_{\Pi} \gamma$ and $\gamma \vdash_{\Pi} \delta$, then $\delta \prec_{\Pi}^{e} \gamma$ and $\gamma \prec_{\Pi}^{e} \delta$. Thus $\prec_{\Pi}^{e}$ is not transitive. For instance, in the concrete example given in $\S 2.1$ we have that $s \forall_{\Pi} r$ and $r \forall_{\Pi} s$, therefore $r \not_{\Pi} s$, $s \nprec_{\Pi} r$ but $r \prec_{\Pi}^{e} s$ and $s \prec_{\Pi}^{e} r$.

Finally we have the relations $\nabla_{p}$ and $\triangleright_{l i t}$ defined, respectively, in terms of the preference relations $\prec_{p}$ and $\prec_{l i t}$. It is clear that both $\prec_{p}$ and $\prec_{l i t}$ are (smooth) orders and the former is also modular. Thus $\triangleright_{p}$ and $\triangleright_{l i t}$ are E-cumulative (by 5.10 ) and in addition $\triangleright_{p}$ satisfies E-R-Cut (by A.10). The associated essential relation of $\triangleright_{p}$ is different than $\prec_{p}$ (for instance, $a \wedge b \prec_{p}^{e} a \wedge \neg b$ ).

## 7 Summary and final comments

The following tables summarize our results. Table 2 shows the properties satisfied by $\nabla_{\swarrow}$ for a given preference relation $\prec$ and Table 3 shows the properties satisfied by the essential relation associated to an explanatory relation (in particular, under those conditions $\triangleright$ is represented by $\prec^{e}$ ).

| $\prec$ |  | $\triangleright_{<}$ |
| :---: | :---: | :---: |
| Smooth | $\Rightarrow$ | E-cumulative |
| Disjunctively smooth | $\Rightarrow$ | E-cumulative + LOR |
| Smooth $+\mathrm{E}_{1}+\mathrm{E}_{2}+$ finite language | $\Rightarrow$ | E-cumulative + RS+ROR |
| Strongly filtered | $\Rightarrow$ | E-cumulative + E-DR |
| Smooth + modular | $\Rightarrow$ | E-cumulative + E-R-Cut |

Table 2: Properties of the explanatory relation associated to a preference relation

We analyzed the connection between structural properties for explanatory relations and preference relations. This detailed analysis gives positive evidence about the correctness of the proposed postulates. In addition, they shed new light on the nature of the postulates. (i) It is reasonable to take E-CM and E-C-Cut as basic rules of explanatory reasoning based on a preference criterion. (ii) Cut rules for observations are essentially connected to preference criteria. (iii) The presence of RS makes the correspondence between explanatory relations and preference relations very tight. However, the examples also suggest that a typical explanatory relation does not satisfies RS.

We presented a variety of examples ranging from the structurally rich causal relations to those, like $\triangleright_{l i t}$, satisfying only the strictly basic E-CM and E-C-Cut. The simplicity criterion of comparing

| E-cumulative $\triangleright$ |  | $\prec^{e}$ |
| :---: | :---: | :---: |
| LOR+ logically finite | $\Rightarrow$ | Disjunctively smooth |
| Full reflexivity | $\Rightarrow$ | Filtered |
| E-R-Cut | $\Rightarrow$ | Filtered |
| RS+ROR+ LOR+ finite language | $\Rightarrow$ | Disjunctively smooth $+\mathrm{E}_{1}+\mathrm{E}_{2}$ |
| E-DR+ RS+ finite language | $\Rightarrow$ | Strongly filtered $+\mathrm{E}_{1}+$ transitive |
| E-DR+E-R-Cut | $\Rightarrow$ | Smooth + modular |

Table 3: Properties of the essential relation associated to an E-cumulative explanatory relation
sets of literals, used for $\triangleright_{l i t}$, is easy to state and compute. But this stands in contrast with the apparently meager logical structure of its associated explanatory relation $\triangleright_{l i t}$. This is an interesting phenomenon which deserves further analysis. It might be the case that explanatory reasoning based on a purely syntactic criterion possess some structural properties we have not considered yet.

From a strict formal point of view, the combinatorial content of the results presented in this paper look very similar to the well known semantical representation theorems for consequence relations [13, 18, 19, 27]. However, it is unclear whether our results can be considered as providing a semantics of some sort for explanatory relations. Thus a very natural question is to determine if there is a semantics for explanatory relations in more traditional terms. The crucial problem is to find a uniform model-based method for representing explanatory relations not satisfying RS. Which, in turn, might lead to isolate other structural rules of explanatory reasoning.

## A Proofs

In order to prove theorem 5.1 we need the following
Lemma A.1. Let $\triangleright$ be an E-cumulative explanatory relation satisfying LOR. Let $\alpha$ be consistent with $\Sigma$ and $\delta \in \operatorname{Expla}(\alpha)$ be an admissible formula such that $\alpha \not \downarrow \delta$. Let

$$
C_{\alpha}=\bigcap\{C n(\Sigma \cup\{\rho\}): \alpha \triangleright \rho\}
$$

and

$$
S=C_{\alpha} \cup\{\neg \beta: \beta \triangleright \delta\} .
$$

Then $S$ is consistent.
Proof. Suppose towards a contradiction, that $S$ is inconsistent. By compactness there are $\beta_{i}$ 's for $i=1, \ldots, n$ such that $\beta_{i} \triangleright \delta$ and $\left(\beta_{1} \vee \cdots \vee \beta_{n}\right) \in C_{\alpha}$. Let $\beta=\beta_{1} \vee \cdots \vee \beta_{n}$. By LOR we know that $\beta \triangleright \delta$. By E-CM we have that $(\alpha \wedge \beta) \triangleright \delta$. Since $\beta \in C_{\alpha}$, then by E-C-Cut we conclude $\alpha \triangleright \delta$, which is a contradiction. Therefore $S$ is consistent.

Proof of theorem 5.1. (ii) $\Rightarrow$ (i) follows from 5.10 and the fact that (12) easily implies LOR. For (i) $\Rightarrow$ (ii), we have already mentioned that the direction from left to right of (14) always holds. On the other hand, if we show that $\prec^{e}$ is smooth and represents $\triangleright$, then from LOR we easily get that $\prec^{e}$ is disjunctively smooth. Thus, we show now that $\prec^{e}$ is smooth and the other direction in (14). Fix
a formula $\alpha$ consistent with $\Sigma$ and let $\delta$ be any formula in $\operatorname{Expla}(\alpha)$. We will show that if $\alpha \not \subset \delta$, then there is $\gamma$ such that $\alpha \triangleright \gamma$ and $\gamma \prec^{e} \delta$. In particular, this will prove that $\prec^{e}$ is smooth and also that the other direction in (14) holds. Suppose $\alpha \not \downarrow \delta$. If $\delta$ is not admissible, then by E-Con there is $\gamma$ such that $\alpha \triangleright \gamma$ and by the definition of $\prec^{e}$ we have $\gamma \prec^{e} \delta$. Hence we will assume that $\delta$ is admissible. Let $C_{\alpha}$ and $S$ be defined as in Lemma A.1. Since $\triangleright$ is logically finite there are two cases to be considered:
(a) Suppose $\triangleright$ satisfies RLF. Let $A=\{\gamma: \alpha \triangleright \gamma\}$ and $\gamma_{i} \in A, i \leq n$ be a lower bound for $A$. It is easy to check that

$$
\begin{aligned}
C_{\alpha} & =\bigcap\left\{C n\left(\Sigma \cup\left\{\gamma_{i}\right\}\right): i \leq n\right\} \\
& =C n\left(\Sigma \cup\left\{\left(\gamma_{1} \vee \cdots \vee \gamma_{n}\right)\right\}\right)
\end{aligned}
$$

Let $N$ be a model of $S$, then there is $i$ such that $N \models \Sigma \cup\left\{\gamma_{i}\right\}$. As $N$ is also a model of $\{\neg \beta: \beta \triangleright \delta\}$, then it is clear that $\gamma_{i} \prec^{e} \delta$.
(b) Suppose $\triangleright$ satisfies LLF. Then the set $\{\beta: \beta \triangleright \delta\}$ has a lower bound. Let $\beta_{1}, \ldots, \beta_{n}$ be such that $\beta_{i} \triangleright \delta$ and $\beta \vdash_{\Sigma} \beta_{1} \vee \cdots \vee \beta_{n}$ for every $\beta$ such that $\beta \triangleright \delta$. Let $\beta^{\prime}=\beta_{1} \vee \cdots \vee \beta_{n}$, then by LOR $\neg \beta^{\prime} \in S$. Since $S$ is consistent then $\beta^{\prime} \notin C_{\alpha}$, hence there is $\gamma$ such that $\alpha \triangleright \gamma$ and $\gamma \forall_{\Sigma} \beta^{\prime}$. Therefore $\gamma \vdash_{\Sigma} \beta$, for all $\beta$ such that $\beta \triangleright \delta$, i.e. $\gamma \prec^{e} \delta$.

In order to prove the Theorem 5.2 we need to describe $\prec^{e}$ in a different way (a similar idea was used in [22, 27]).

Definition A.2. Let $\triangleright$ be an explanatory relation. Define a binary relation $\prec^{u}$ as follows. Let $\gamma$ be an admissible formula

$$
\gamma \prec^{u} \delta \stackrel{\text { def }}{\Leftrightarrow} \forall \alpha \forall \beta[\alpha \triangleright \gamma \& \beta \triangleright \delta \Rightarrow(\alpha \vee \beta) \triangleright \gamma \&(\alpha \vee \beta) \ngtr \delta]
$$

Lemma A.3. Let $\triangleright$ be an E-cumulative explanatory relation satisfying E-DR. Then $\prec^{e}=\prec^{u}$. Moreover, $\prec^{u}$ (and therefore $\prec^{e}$ ) is transitive.
Proof. $\left(\prec^{e} \subseteq \prec^{u}\right)$ : Let $\gamma \prec^{e} \delta$ and $\alpha \triangleright \gamma, \beta \triangleright \delta$. By the definition of $\prec^{e}$ we have $(\alpha \vee \beta) \ngtr \delta$ and by E-DR we have $(\alpha \vee \beta) \triangleright \gamma$. Hence $\gamma \prec^{u} \delta$.
$\left(\prec^{u} \subseteq \prec^{e}\right)$ : Let $\gamma, \delta$ be admissible formulas with $\gamma \prec^{u} \delta$. Suppose, towards a contradiction, that there is $\beta$ such that $\beta \triangleright \delta$ and $\gamma \vdash_{\Sigma} \beta$. Let $\alpha$ be any formula such that $\alpha \triangleright \gamma$. Since $\gamma \vdash_{\Sigma} \beta$, then by E-CM we have $(\alpha \wedge \beta) \triangleright \gamma$. Since $\vdash((\alpha \wedge \beta) \vee \beta) \leftrightarrow \beta$ and $\gamma \prec^{u} \delta$, then (by LLE $_{\Sigma}$ ) we conclude that $\beta \not \downarrow \delta$, which is a contradiction.

To see that $\prec^{u}$ is transitive, let $\gamma_{i}$ be formulas such that $\gamma_{1} \prec^{u} \gamma_{2}$ and $\gamma_{2} \prec^{u} \gamma_{3}$. Without lost of generality we can assume that each $\gamma_{i}$ is admissible. Let $\alpha_{i}$ be formulas such that $\alpha_{i} \triangleright \gamma_{i}$. By E-DR it suffices to show that $\left(\alpha_{1} \vee \alpha_{3}\right) \not \gamma_{3}$. Suppose, towards a contradiction, that $\left(\alpha_{1} \vee \alpha_{3}\right) \triangleright \gamma_{3}$. Since $\gamma_{2} \prec^{u} \gamma_{3}$, then by definition of $\prec^{u}$ we have $\left(\alpha_{1} \vee \alpha_{2} \vee \alpha_{3}\right) \triangleright \gamma_{2}$ and $\left(\alpha_{1} \vee \alpha_{2} \vee \alpha_{3}\right) \not \gamma_{3}$. Since $\gamma_{1} \prec^{u} \gamma_{2}$, then analogously we have $\left(\alpha_{1} \vee \alpha_{2} \vee \alpha_{3}\right) \triangleright \gamma_{1}$ and $\left(\alpha_{1} \vee \alpha_{2} \vee \alpha_{3}\right) \ngtr \gamma_{2}$, which is a contradiction.

Remark: Without the hypothesis that E-DR holds the previous result fails. For instance, in example $\S 6.5$ it is shown that $\prec_{\Pi}^{e}$ is not transitive, even though $\triangleright_{\Pi}$ is E-cumulative (and moreover satisfies E-R-Cut).

Lemma A.4. Let $\triangleright$ be a E-cumulative explanatory relation satisfying also LLF, RS and E-DR. Then $\prec^{u}$ (and therefore $\prec^{e}$ ) is strongly filtered.

Proof. Let $\alpha$ be a formula and $\gamma, \delta \in \operatorname{Expla}(\alpha)$ be such that $\gamma, \delta \notin \min \left(\operatorname{Expla}(\alpha), \prec^{e}\right)$. Consider $A=\{\beta: \beta \triangleright \gamma\}$ and $B=\{\theta: \theta \triangleright \delta\}$. By LLF and LOR (which follows from E-DR), there is $\beta_{0} \in A$ and $\theta_{0} \in B$ such that $\beta \vdash \beta_{0}$ and $\theta \vdash \theta_{0}$ for all $\beta \in A$ and $\theta \in B$. By E-DR, either $\left(\beta_{0} \vee \theta_{0}\right) \triangleright \gamma$ or $\left(\beta_{0} \vee \theta_{0}\right) \triangleright \delta$. By E-CM either $\left(\beta_{0} \vee \theta_{0}\right) \wedge \alpha \triangleright \gamma$ or $\left(\beta_{0} \vee \theta_{0}\right) \wedge \alpha \triangleright \delta$. Since $\alpha \not{ }^{2}$, then in either case by E-C-Cut there is $\rho_{0}$ such that $\alpha \triangleright \rho_{0}$ and $\rho_{0} \nvdash_{\Sigma}\left(\beta_{0} \vee \theta_{0}\right)$. Let $\rho=\rho_{0} \wedge \neg \beta_{0} \wedge \neg \theta_{0}$. Then by $\mathrm{RS}, \alpha \triangleright \rho$ and if $\rho^{\prime} \vdash \rho$, then $\rho^{\prime} \vdash \neg \beta_{0} \wedge \neg \theta_{0}$ and thus $\rho^{\prime} \prec^{e} \gamma$ and $\rho^{\prime} \prec^{e} \delta$.

Lemma A.5. If $\prec$ is a strongly filtered preference relation, then $\nabla^{2}$ satisfies E-DR.
Proof. Let $\triangleright=\triangleright_{\alpha}$. Let $\alpha \triangleright \gamma$ and $\beta \triangleright \delta$. Suppose, towards a contradiction, that $(\alpha \vee \beta) \ngtr \gamma$ and $(\alpha \vee \beta) \nvdash \delta$. Therefore $\gamma, \delta \notin \min (\operatorname{Expla}(\alpha \vee \beta), \prec)$. As $\prec$ is strongly filtered, there is $\rho$ such that $(\alpha \vee \beta) \triangleright \rho, \rho \prec \gamma, \rho \prec \delta$ and $\rho^{\prime} \prec \gamma$ and $\rho^{\prime} \prec \delta$ for every $\rho^{\prime} \vdash \rho$. Since $\rho \prec \gamma$, then $\rho \vdash_{\Sigma} \alpha$. Thus there is $\rho^{\prime}$ consistent with $\Sigma$ such that $\rho^{\prime} \vdash \rho$ and $\rho^{\prime} \vdash_{\Sigma} \neg \alpha$. Therefore $\rho^{\prime} \vdash_{\Sigma} \beta$, which contradicts that $\rho^{\prime} \prec \delta$.

Lemma A.6. Let $\triangleright$ be an explanatory relation satisfying RS. Then $\prec^{e}$ satisfies $\mathrm{E}_{1}$.
Proof. Suppose that $\delta \prec^{e} \gamma$. Thus by definition $\delta$ is admissible. If $\gamma^{\prime}$ is not admissible then by definition $\delta \prec^{e} \gamma^{\prime}$. Now suppose that $\gamma^{\prime}$ is admissible and also, towards a contradiction, that $\delta \not \kappa^{e} \gamma^{\prime}$. Let $\beta$ be such that $\beta \triangleright \gamma^{\prime}$ and $\delta \vdash_{\Sigma} \beta$. Since $\gamma \vdash_{\Sigma} \gamma^{\prime}$ and $\gamma \nvdash_{\Sigma} \perp$, we have by RS that $\beta \triangleright \gamma$ and therefore $\delta \kappa^{e} \gamma$.

The last fact we need to establish the Theorem 5.2 is the following straightforward result.
Lemma A.7. If $\prec$ is a preference relation satisfying $\mathrm{E}_{1}$, then $\nabla^{2}$ satisfies RS .
Proof of Theorem 5.2. It is enough to put together 5.1, A.6, A.7, A.3, A. 4 and A.5.
In order to prove Theorem 5.3, it is convenient to show first 5.10 and 5.13.
Proof of Theorem 5.10. It is straightforward to check that E-Reflexivity and E-CM hold. If $\prec$ is smooth, then easily $\mathrm{E}-\mathrm{Con}_{\Sigma}$ holds. To see E-C-Cut, suppose that the premises in the rule E-C-Cut hold. Hence $\min (\operatorname{Expla}(\alpha)) \subseteq \operatorname{Expla}(\beta)$ and since $\operatorname{Expla}(\alpha \wedge \beta) \subseteq \operatorname{Expla}(\alpha)$, then $\min (\operatorname{Expla}(\alpha)) \subseteq \min (\operatorname{Expla}(\alpha \wedge \beta))$. Since $\operatorname{Expla}(\alpha)$ is smooth we conclude $\min (\operatorname{Expla}(\alpha))=$ $\min (\operatorname{Expla}(\alpha \wedge \beta))$ and this finishes the proof.

Proof of the proposition 5.13. As we have argued before, it suffices to show that if $\alpha \nvdash_{\Sigma} \perp, \gamma \vdash_{\Sigma} \alpha$ and $\alpha \not \downarrow \gamma$, then there is $\delta \prec^{e} \gamma$ with $\alpha \triangleright \delta$. By E-Con ${ }_{\Sigma}$ there is $\delta$ such that $\alpha \triangleright \delta$. We will show that $\delta \prec^{e} \gamma$. Suppose otherwise, then there is $\beta$ such that $\delta \vdash_{\Sigma} \beta$ and $\beta \triangleright \gamma$. By E-CM, $(\alpha \wedge \beta) \triangleright \gamma$ and by E-R-Cut, $\alpha \triangleright \gamma$ which is a contradiction. The previous argument also shows that $\prec^{e}$ is filtered.

Now we establish some other auxiliary results.
Lemma A.8. Let $\triangleright$ be an E-cumulative explanatory relation satisfying also E-DR and E-R-Cut. Then for all admissible formulas $\gamma$ and $\delta$,

$$
\begin{equation*}
\gamma \prec^{u} \delta \Leftrightarrow \exists \alpha \exists \beta[\alpha \triangleright \gamma \& \beta \triangleright \delta \&(\alpha \vee \beta) \triangleright \gamma \&(\alpha \vee \beta) \ngtr \delta] \tag{20}
\end{equation*}
$$

Proof. The $(\Rightarrow)$ direction comes directly from the definition of $\prec^{u}$. For the other direction, let $\alpha$ and $\beta$ be as in the right hand side of (20) and $\alpha^{\prime}$ and $\beta^{\prime}$ be formulas such that $\alpha^{\prime} \triangleright \gamma$ and $\beta^{\prime} \triangleright \delta$. We need to show that $\left(\alpha^{\prime} \vee \beta^{\prime}\right) \triangleright \gamma$ and $\left(\alpha^{\prime} \vee \beta^{\prime}\right) \nvdash \delta$. Since $\triangleright$ satisfies E-DR, hence it suffices to show that $\left(\alpha^{\prime} \vee \beta^{\prime}\right) \not$. $^{\prime}$. Suppose, towards a contradiction, that $\left(\alpha^{\prime} \vee \beta^{\prime}\right) \triangleright \delta$. By E-CM we have $\left(\alpha^{\prime} \vee \beta^{\prime}\right) \wedge(\alpha \vee \beta) \triangleright \delta$. And by hypothesis $(\alpha \vee \beta) \triangleright \gamma$ and clearly $\gamma \vdash_{\Sigma}\left(\alpha^{\prime} \vee \beta^{\prime}\right)$, hence by E-R-Cut $(\alpha \vee \beta) \triangleright \delta$, which contradicts the choice of $\alpha$ and $\beta$.

We will show next that when $\triangleright$ satisfies E-DR and E-R-Cut then $\prec^{u}$ is modular.
Lemma A.9. Let $\triangleright$ be an E-cumulative explanatory relation satisfying E-DR and E-R-Cut. Then $\prec^{e}$ is a smooth and modular preference relation that represents $\triangleright$. If in addition, $\triangleright$ satisfies RS , then $\prec^{e}$ is strongly filtered.

Proof. For the first claim, by proposition 5.13 it remains to be shown that $\prec^{e}$ (alias $\prec^{u}$, by A.3) is modular. Let $\gamma, \delta$ and $\rho$ be formulas such that $\gamma \nprec^{u} \delta, \delta \nprec^{u} \gamma$ and $\gamma \prec^{u} \rho$. We want to show that $\delta \prec^{u} \rho$. Without lost of generality we can assume that $\gamma, \delta$ and $\rho$ are admissible. Let $\alpha, \beta$, $\omega$ formulas such that $\alpha \triangleright \gamma, \beta \triangleright \delta$ and $\omega \triangleright \rho$. Since $\gamma$ and $\delta$ are $\prec^{u}$-incomparable then from (20) it follows that $(\alpha \vee \beta) \triangleright \gamma$ and $(\alpha \vee \beta) \triangleright \delta$. Again by (20) it suffices to show that $(\alpha \vee \beta \vee \omega) \triangleright \delta$
 definition of $\prec^{u}$ we have $(\alpha \vee \beta \vee \omega) \triangleright \gamma$ and $(\alpha \vee \beta \vee \omega) \ngtr \rho$. Finally, suppose that $\triangleright$ satisfies RS. An argument analogous to that in the proof of proposition 5.13 shows that $\prec^{e}$ is strongly filtered.

Lemma A.10. If $\prec$ is a smooth and modular preference relation, then $D_{\prec}$ is $E$-cumulative and satisfies E-R-Cut.

Proof. Let $\triangleright=\triangleright_{\alpha}$. From theorem 5.10 we know that $\triangleright$ is E-cumulative. It remains to be shown that E-R-Cut holds. Let $\alpha, \beta, \gamma$ and $\delta$ formulas such that $(\alpha \wedge \beta) \triangleright \gamma, \alpha \triangleright \delta$ and $\delta \vdash_{\Sigma} \beta$. We need to show that $\alpha \triangleright \gamma$. Suppose, towards a contradiction, that $\alpha \not \downarrow \gamma$. Since $\gamma \vdash_{\Sigma} \alpha$, then by smoothness and the definition of $\triangleright$, there is $\delta^{\prime}$ such that $\alpha \triangleright \delta^{\prime}$ and $\delta^{\prime} \prec \gamma$. Since $\alpha \triangleright \delta$ then $\delta \nprec \delta^{\prime}$ and $\delta^{\prime} \nprec \delta$. By modularity, $\delta \prec \gamma$ and by E-CM $(\alpha \wedge \beta) \triangleright \delta$, which contradicts the hypothesis that $(\alpha \wedge \beta) \triangleright \gamma$.

Remark: One would expect to have also E-DR in the conclusion of A.10, but this is not true. For instance, the preference relation $\prec_{p}$ (see $\S 2.2$ ) is modular and smooth but $\triangleright_{p}$ does not even satisfy LOR.

Proof of the Theorem 5.3. We recall that any rational relation satisfies E-DR (see [26, 2.18]), it is enough to put together A.6, A.7, A. 9 and A. 10.

Now we are going to prove the Theorem 5.4. In order to do that we need to recall some results. Since any E-cumulative causal explanatory relation necessarily satisfies RS, then its associated consequence relation $\sim_{a b}$ is preferential [26, theorem 2.8] and therefore it is represented by a preferential model [18]. We will use this important result in the proof of 5.4. We recall below the notion of a preferential model.

Definition A.11. A structure $\mathcal{M}$ is a triple $\langle S, \imath, \sqsubset\rangle$ where $S$ is a set (called the set of states), $\sqsubset$ is a strict order (i.e. transitive and irreflexive) on $S$ and $\imath: S \longrightarrow \mathcal{U}$ is a function (called the interpretation function).

Let $\mathcal{M}=\langle\mathcal{S}, \imath, \sqsubset\rangle$ be a structure. We adopt the following notations: if $T \subseteq S$, then $\min (T)=$ $\left\{t \in T: \nexists t^{\prime} \in T, t^{\prime} \sqsubset t\right\}$, i.e. $\min (T)$ is the set of all minimal elements of $T$ with respect to $\sqsubset$; $\bmod _{\mathcal{M}}(\alpha)=\{s \in S: \imath(s) \models \alpha\} ; \min _{\mathcal{M}}(\alpha)=\min \left(\bmod _{\mathcal{M}}(\alpha)\right)$.

Definition A.12. Let $\mathcal{M}=\langle\mathcal{S}, \imath, \sqsubset\rangle$ be a structure and $T \subseteq S$. We say that $T$ is smooth if it satisfies the following

$$
\forall s \in T \backslash \min (T) \quad \exists s^{\prime} \in \min (T) s^{\prime} \sqsubset s
$$

$\mathcal{M}$ is said to be a preferential model if $\bmod _{\mathcal{M}}(\alpha)$ is smooth for any formula $\alpha$.
Each preferential model has an associated consequence relation given by the following:
Definition A.13. Let $\mathcal{M}=\langle\mathcal{S}, \imath, \sqsubset\rangle$ be a preferential model. The consequence relation $\sim_{\mathcal{M}}$ is defined by the following

$$
\begin{equation*}
\alpha \sim_{\mathcal{M}} \beta \Leftrightarrow \min _{\mathcal{M}}(\alpha) \subseteq \bmod _{\mathcal{M}}(\beta) \tag{21}
\end{equation*}
$$

and the nonmonotonic consequence of a formula is given by

$$
C_{\mathcal{M}}(\alpha)=\left\{\beta: \alpha \sim_{\mathcal{M}} \beta\right\}
$$

The following representation theorem is one of the basic tools in the study of nonmonotonic consequence relations.

Theorem A.14. (Kraus, Lehmann and Magidor [18]) A consequence relation $\sim$ is preferential if and only if there is a preferential model $\mathcal{M}$ such that $\sim=\sim_{\mathcal{M}}$.

Now we can use theorem A. 14 to study the properties of $\prec^{e}$. We start with a simple but useful fact concerning the admissible formulas.

Lemma A.15. A formula $\gamma$ is admissible if and only if $\bmod (\{\gamma\} \cup \Sigma)=\left\{\imath(s): s \in \min _{\mathcal{M}}(\gamma)\right\}$.
Proof. $\gamma$ is admissible iff $\bmod (\{\gamma\} \cup \Sigma) \subseteq \bmod (C(\gamma))$ iff $\bmod (\{\gamma\} \cup \Sigma)=\left\{\imath(s): s \in \min _{\mathcal{M}}(\gamma)\right\}$.

Lemma A.16. Suppose the language is finite. Let $\triangleright$ be a causal E-cumulative explanatory relation. Let $\mathcal{M}=\langle\mathcal{S}, \imath, \sqsubset\rangle$ be a preferential model for $\sim_{a b}$ given by theorem A.14. Then for $\gamma$ and $\delta$ admissible

$$
\begin{equation*}
\gamma \prec^{e} \delta \Rightarrow \exists s \in \min _{\mathcal{M}}(\gamma) \exists t \in \min _{\mathcal{M}}(\delta) s \sqsubset t \tag{22}
\end{equation*}
$$

Moreover, if $\imath$ is injective, then in (22) the equivalence holds.
Proof. Suppose that $s \not \subset t$ for all $s \in \min _{\mathcal{M}}(\gamma)$ and all $t \in \min _{\mathcal{M}}(\delta)$. We will show that $(\gamma \vee \delta) \triangleright \delta$, which clearly implies that $\gamma \not_{e} \delta$. By lemma A.15, it is enough to see that $\min _{\mathcal{M}}(\delta) \subseteq \min _{\mathcal{M}}(\gamma \vee \delta)$. Let $t$ be in $\min _{\mathcal{M}}(\delta)$ and suppose towards a contradiction that $t \notin \min _{\mathcal{M}}(\gamma \vee \delta)$. It is clear that $t \in \bmod _{\mathcal{M}}(\gamma \vee \delta)$ and by smoothness of $\sqsubset$ there is $s \in \min _{\mathcal{M}}(\gamma \vee \delta)$ and $s \sqsubset t$. Since $t \in \min _{\mathcal{M}}(\delta)$, then necessarily $s \in \min _{\mathcal{M}}(\gamma)$, which is a contradiction.

Suppose now that $\imath$ is injective. To show $(\Leftarrow)$ in $(22)$, let $\alpha \triangleright \delta, s \in \min _{\mathcal{M}}(\gamma)$ and $t \in \min _{\mathcal{M}}(\delta)$ such that $s \sqsubset t$. Notice that $\imath(t)$ is a model of $C_{\mathcal{M}}(\alpha)$ and since $\imath$ is injective (and the language is finite), then necessarily $t \in \min _{\mathcal{M}}(\alpha)$. Then $\imath(s) \not \vDash \alpha$, as $s \sqsubset t$. Therefore $\gamma \prec^{e} \delta$.

Lemma A.17. Let $\triangleright$ be an E-cumulative explanatory relation.
(i) If in addition $\triangleright$ satisfies RLF and ROR, then $\prec^{e}$ is filtered.
(ii) Suppose the language is finite. If in addition $\triangleright$ satisfies LOR, RS and ROR, then $\prec^{e}$ satisfies $\mathrm{E}_{2}$.

Proof. (i) To see that $\prec^{e}$ is filtered, let $\alpha$ be a formula and $\gamma, \delta \in \operatorname{Expla}(\alpha)$ be such that $\gamma, \delta \notin$ $\min \left(\operatorname{Expla}(\alpha), \prec^{e}\right)$. Consider $\{\rho: \alpha \triangleright \rho\}$. By RLF, let $\rho_{1}, \cdots, \rho_{n}$ be a lower bound. We claim that $\rho=\rho_{1} \vee \cdots \vee \rho_{n}$ works. In fact, by ROR, $\alpha \triangleright \rho$ and thus $\rho \in \min \left(\operatorname{Expla}(\alpha), \prec^{e}\right)$. Now we will show that $\rho \prec^{e} \gamma$. Suppose not and let $\beta$ be such that $\rho \vdash_{\Sigma} \beta$ and $\beta \triangleright \gamma$. By E-CM, $(\alpha \wedge \beta) \triangleright \gamma$. Since $\rho_{i} \vdash \beta$ for all $i$, then by E-C-Cut $\alpha \triangleright \gamma$, which is a contradiction. Therefore $\rho \prec^{e} \gamma$. Analogously it is shown that $\rho \prec^{e} \delta$.
(ii) We recall that (for a finite language) if $\triangleright$ is E-preferential and satisfies LOR, then $\vdash_{a b}$ admits an injective preferential model $\mathcal{M}=\langle\mathcal{S}, \imath, \sqsubset\rangle$ (by [26, theorem 2.9] and [13, theorem 4.13]). We will use A. 16 to show the contrapositive of $\mathrm{E}_{2}$. Suppose $\rho \kappa^{e} \gamma$ and $\rho \kappa^{e} \delta$. To see that $\rho \nprec^{e} \gamma \vee \delta$, let $t \in \min _{\mathcal{M}}(\gamma \vee \delta)$ and $s \in \min _{\mathcal{M}}(\rho)$. It is clear that $t \in \min _{\mathcal{M}}(\gamma) \cup \min _{\mathcal{M}}(\delta)$. Then by A. 16 it is clear that $s \not \subset t$.

The following is a straightforward result.
Lemma A.18. If $\prec$ is a preference relation satisfying $\mathrm{E}_{2}$, then $\nabla^{2}$ satisfies ROR.

Finally we have
Proof of theorem 5.4. (i) We will show that $\triangleright$ is represented by $\prec^{e}$. Let $\mathcal{M}=\langle\mathcal{S}, \imath, \prec\rangle$ be a preferential model for $\mu_{a b}$ given by theorem A.14. As in the proof of theorem 5.1, we have only to show that if $\delta$ is an admissible formula such that $\delta \vdash_{\Sigma} \alpha$ and $\alpha \nvdash \delta$, then there is $\gamma \prec^{e} \delta$ with $\alpha \triangleright \gamma$. Suppose then that $\delta$ is admissible and $\alpha \not \downarrow \delta$. Consider the following sets

$$
A=\left\{t \in S: t \in \min _{\mathcal{M}}(\delta) \backslash \min _{\mathcal{M}}(\alpha)\right\}
$$

and

$$
B=\left\{s \in S: s \in \min _{\mathcal{M}}(\alpha) \& s \sqsubset t \text { for some } t \in A\right\}
$$

From (11) and the fact that $\mathcal{M}$ represents $\vdash_{a b}$, we conclude that there is $t \in \bmod _{\mathcal{M}}(\delta)$ such that $t \notin \min _{\mathcal{M}}(\alpha)$. Moreover, by lemma A. 15 we know that $t \in \min _{\mathcal{M}}(\delta)$. By smoothness of $\sqsubset$, there is $s \in \min _{\mathcal{M}}(\alpha)$ with $s \sqsubset t$. Thus $B \neq \emptyset$. Since the language is finite, let $\gamma$ be a formula such that

$$
\bmod (\gamma)=\{\imath(u): u \in B\}
$$

It is straightforward to verify that $\alpha \triangleright \gamma$. We claim that $\gamma \prec^{e} \delta$. Suppose, towards a contradiction, that there is $\beta$ such that $\beta \triangleright \delta$ and $\gamma \vdash_{\Sigma} \beta$. Fix $N \models \delta$. Since $\beta \triangleright \delta$, then $N \models C(\beta)$, therefore there is $u \in \min _{\mathcal{M}}(\beta)$ such that $\imath(u)=N$. It is clear that $u \in \min _{\mathcal{M}}(\delta)$ and $u \notin A$. Thus $u \in \min _{\mathcal{M}}(\alpha)$, hence $N \models C(\alpha)$. Therefore $\alpha \triangleright \delta$ which is a contradiction.
(ii) It follows from 5.1, A.6, A.7, A. 17 and A. 18.

Proof of the Proposition 5.12. Let us start by checking that $\triangleright$ is represented by $\prec^{e}$. Since the direction from left to right of (14) always holds, it suffices to show that if $\alpha \nvdash_{\Sigma} \perp, \gamma \vdash_{\Sigma} \alpha$ and $\alpha \not{ }^{\prime}$
then there is $\delta \prec^{e} \gamma$ with $\alpha \triangleright \delta$. We claim that $\delta=\alpha$ works. In fact, by full reflexivity we have that $\alpha \triangleright \alpha$. Let us assume, towards a contradiction, that there is $\beta$ such that $\alpha \vdash_{\Sigma} \beta$ and $\beta \triangleright \gamma$. Then, by E-CM, $(\alpha \wedge \beta) \triangleright \gamma$ and, by $\operatorname{LLE}_{\Sigma}, \alpha \triangleright \gamma$ which is a contradiction. Thus $\alpha \prec^{e} \gamma$. The same argument shows that $\prec^{e}$ is filtered.

We give now two examples. The first one shows that the essential preference relation $\prec^{e}$ associated with a causal explanatory relation does not need to satisfy $\mathrm{E}_{2}$, so the hypothesis LOR in A.17(ii) is necessary. The second one shows that there are E-cumulative relations which are not representable by its associated preference relation.

Example A.19. The language will be finite with at least two atoms. Consider the following preferential model $\mathcal{M}=\langle\mathcal{S}, \imath, \sqsubset\rangle$ where $S=\{a, b, c, d\}$ and $\sqsubset$ is given by $a \sqsubset b$ and $c \sqsubset d$. Let $P, N$ and $M$ be three different valuations. The function $\imath$ is given by $\imath(a)=N, \imath(c)=M$ and $\imath(b)=\imath(d)=P$. Let $\sim_{\mathcal{M}}$ be the consequence relation associated with $\mathcal{M}$. Let $\tilde{\triangleright}$ be defined by $\alpha \tilde{\triangleright} \gamma$ iff $C_{\mathcal{M}}(\alpha) \subseteq C n(\gamma)$. It is routine to check that $\tilde{\triangleright}$ is a E-preferential causal explanatory relation. Let $\prec_{c}^{e}$ denote the essential preference relation associated with $\tilde{\triangleright}$. We will show that $\prec_{c}^{e}$ does not satisfy $\mathrm{E}_{2}$. Let $\gamma, \delta$ and $\rho$ be formulas whose models are, respectively, $\{N, P\},\{M, P\}$ and $\{P\}$. It is clear that $\gamma \tilde{\triangleright} \gamma, \delta \tilde{\triangleright} \delta$ and $\rho \tilde{\triangleright} \rho$. Notice that $\gamma \vee \delta$ is not admissible, in fact the models of $C_{\mathcal{M}}(\gamma \vee \delta)$ are $N$ and $M$. Thus $\rho \prec_{c}^{e}(\gamma \vee \delta)$ but $\rho \not_{c}^{e} \gamma$ and $\rho \not_{c}^{e} \delta$ as $\rho \vdash(\gamma \wedge \delta)$.

We will slightly modify $\tilde{\triangleright}$ as follows: define $\alpha \triangleright \gamma$ if $\alpha \tilde{\triangleright} \gamma$ and $\gamma$ has only one model. It is routine to verify that $\triangleright$ is E-cumulative and satisfies RS but not ROR. We will show $\triangleright$ is not representable by its associated essential preference relation $\prec^{e}$. Let $\gamma, \delta$ and $\rho$ denote the same formulas as above. Then $\gamma \triangleright \rho$ and $\delta \triangleright \rho$. Let $\sigma$ and $\tau$ be formulas such that its only model is, respectively, $N$ and $M$. Then it is clear that $\sigma \nprec^{e} \rho$ and $\tau \nprec e \rho$. Thus $\rho$ is a $\prec^{e}$-minimal element of $\operatorname{Expla}(\top)$. However, it is clear that $\top \not \rho$.

## References

[1] Atocha Aliseda-LLera. Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence. PhD thesis, Stanford University, 1997. http://www.wins.uva.nl/research/illc/wwwdissertations.html.
[2] Atocha Aliseda-LLera. Abduction as epistemic change: a Peircean model in artificial intelligence. In P. Flach and A. Kakas, editors, Abduction and Induction. Kluwer Academic Publishers, 2000.
[3] I. Bloch and J. Lang. Towards Mathematical Morpho-Logics. In 8th International Conference on Information Processing and Management of Uncertainty in Knowledge based Systems IPMU 2000, volume III, pages 1405-1412, Madrid, Spain, july 2000.
[4] I. Bloch, R. Pino-Pérez, and C. Uzcátegui. Explanatory relations based on mathematical morphology. In ECSQARU 2001, pages 736-747, Toulouse, France, September 2001.
[5] C. Boutilier and V. Becher. Abduction as belief revision. Artificial Intelligence, 77:43-94, 1995.
[6] M. Cialdea Mayer and F. Pirri. Abduction is not deduction-in-reverse. Journal of the IGPL, 4(1):1-14, 1996.
[7] D. Dubois and H. Prade. Epistemic entrenchment and possibilistic logic. Artificial Intelligence, 50:223-239, 1991.
[8] M. A. Falappa, G. Kern-Isberner, and G. R. Simari. Explanations, belief revision and defeasible reasoning. Artificial Intelligence, 141:1-28, 2002.
[9] P. A. Flach. Rationality postulates for induction. In Yoav Shoam, editor, Proc. of the Sixth Conference of Theoretical Aspects of Rationality and Knowledge (TARK96), pages 267-281, March 17-20, 1996.
[10] P. Flach and A. Kakas, editors. Abduction and Induction. Kluwer Academic Publishers, 2000.
[11] P. Flach. Logical characterisations of inductive learning. In D. M. Gabbay and R. Kruse, editors, Abductive Reasoning and Learning, pages 155-196. Kluwer Academic Publishers, 2000.
[12] P. Flach. On the logic of hypothesis generation. In P. Flach and A. Kakas, editors, Abduction and Induction, pages 89-106. Kluwer Academic Publishers, 2000.
[13] M. Freund. Injective models and disjunctive relations. J.Logic Computat, 3:231-247, 1993.
[14] D. M. Gabbay and R. Kruse, editors. Abductive Reasoning and learning. Kluwer Academic Publishers, 2000.
[15] P. Gärdenfors and D. Makinson. Nonmonotonic inferences based on expectations. Artificial Intelligence, 65:197-245, 1994.
[16] P. Gärdenfors. Knowledge in Flux: modeling the dynamics of epistemic states. MIT press, Cambridge, MA, 1988.
[17] K. Konolige. Abduction versus closure in causal theories. Artificial Intelligence, 53:255-272, 1992. Research Note.
[18] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. Artificial Intelligence, 44(1):167-207, 1990.
[19] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? Artificial Intelligence, 55:1-60, 1992.
[20] H. J. Levesque. A knowledge level account of abduction. In Proceedings of the eleventh Internattional Joint Conference on Artificial Intelligence, pages 1061-1067, Detroit, 1989.
[21] J. Lobo and C. Uzcátegui. Abductive change operators. Fundamenta Informaticae, 27(4):385411, 1996.
[22] J. Lobo and C. Uzcátegui. Abductive consequence relations. Artificial Intelligence, 89(1-2):149-171, 1997.
[23] D. Makinson. Five faces of minimality. Studia Logica, 52:339-379, 1993.
[24] D. Makinson. General patterns in nonmonotonic reasoning. In C. Hogger D. Gabbay and J. Robinson, editors, Handbook of Logic in Artificial Intelligence and Logic Programming, volume III, Nonmonotonic Reasoning and Uncertain Reasoning. Oxford University Press, 1994.
[25] L. Palopoli, F. Pirri, and C. Pizzuti. Algorithms for selective enumerations of prime implicants. Artificial Intelligence, 111:41-72, 1999.
[26] R. Pino-Pérez and C. Uzcátegui. Jumping to explanation versus jumping to conclusions. Artificial Intelligence, 111(2):131-169, 1999.
[27] R. Pino-Pérez and C. Uzcátegui. On representation theorems for non-monotonic consequence relations. Journal of Symbolic Logic, 65(3):1321-1337, 2000.
[28] R. Pino-Pérez and C. Uzcátegui. Ordering explanations and the structural rules for abduction. In B. Selman A. Cohn, F. Giunchiglia, editor, Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR2000), pages 637-646, 2000.


[^0]:    ${ }^{1}$ This seems to be related to the explanatory power of an inference, a notion attributed to Hempel in [1, 6].
    ${ }^{2}$ For an overall picture of various forms of defeasible reasoning within the general semantic framework of minimality see [23].

[^1]:    ${ }^{3}$ As it is well known, an analogous phenomenon occurs within the KLM formalism for nonmonotonic reasoning and AGM formalism for belief revision.

[^2]:    ${ }^{4}$ We have restricted ourselves to the so called injective preferential models, but similar constructions can be done with arbitrary preferential models [18]. Various methods based on Mathematical Morphology for selecting models of a formula were suggested in [3] and applied to the context of explanatory reasoning in [4].

[^3]:    ${ }^{5}$ Some of these postulates were incorrectly named in [26]. RS was called there Right And and denoted RA. ROR was called Right Weakening and denoted E-RW. In that paper ROR denotes a different postulate, a sort of elimination of $\vee$ in the right.

[^4]:    ${ }^{6}$ It is known that E-CM + E-C-Cut + RS + E-R-Cut implies E-DR (see [26, Proposition 2.14]), this example shows that $\mathrm{E}-\mathrm{CM}+\mathrm{E}-\mathrm{C}-\mathrm{Cut}+\mathrm{E}-$ R-Cut does not suffice.
    ${ }^{7}$ For orders over a finite set $S$, being filtered is equivalent to require that $(S, \prec)$ does not contain an order isomorphic copy of the order on $\{a, b, c, d\}$ given by $a<b$ and $c<d$ (these two pairs only).

[^5]:    ${ }^{8}$ In [26] E-Con ${ }_{\Sigma}$ was not included as part of E-cumulativity, but all explanatory relations found in this paper will satisfy this postulate. On the other hand, $\mathrm{RLE}_{\Sigma}$ was also included in [26] as part of the E-cumulative system, where $\mathrm{RLE}_{\Sigma}$, Right Logical Equivalence, is defined analogously as $\operatorname{LLE}_{\Sigma}$ but for the right hand side. $\mathrm{RLE}_{\Sigma}$ turns out to be unnecessary. For E-preferential relations it does not make any difference, since RS implies RLE $\boldsymbol{E}_{\Sigma}$ [26, Proposition 2.6].
    ${ }^{9}$ In [4] a method to construct explanatory relations which are not E-cumulative was introduced.
    ${ }^{10}$ For a more elaborated analysis of the connection between causal explanatory relations and belief revision see [26, $\S 4]$.

[^6]:    ${ }^{11}$ We should have said upper bound in the usual sense of the lattice of formulas. However, it is customary when dealing with explanations to use the reversed lattice order.
    ${ }^{12}$ In the first version of this paper [28] we also required that preference relations must be invariant under logical equivalence with respect to $\Sigma$, i.e. if $\alpha \prec \beta$ and $\vdash_{\Sigma}\left(\alpha \leftrightarrow \alpha^{\prime}\right)$ and $\vdash_{\Sigma}\left(\beta \leftrightarrow \beta^{\prime}\right)$, then $\alpha^{\prime} \prec \beta^{\prime}$. This restriction turns out to be unnecessary and restrictive. For instance, it will rule out $\prec_{\text {lit }}$.

[^7]:    ${ }^{13}$ This postulate says that if $\alpha \vee \beta \nsim \rho$, then $\alpha \sim \rho$ or $\beta \sim \rho$.

[^8]:    ${ }^{14}$ That is to say, the order $<$ defined by $\gamma<\delta$ if $\neg \delta \prec_{n c}^{e} \neg \gamma$ is an expectation order.

