Universidad de los Andes

Facultad de Ciencias

Departamento de Matemática

Ordering explanations and the structural rules for abduction

Ramón Pino-Pérez and Carlos Uzcátegui

Notas de Matemática

Serie: Pre-Print

No. 202

Mérida - Venezuela

2000

Notas de Matemática, No. 202
Mérida, 2000.

Ordering explanations and the structural rules for abduction

Ramón Pino-Pérez and Carlos Uzcátegui

Abstract

We study the relationship between some structural rules for abductive reasoning and preference relations for selection preferred explanations. We prove that explanatory relations having good structural properties can always be defined by orders over formulas.

1 Introduction

Abduction is usually defined as the process of inferring the best explanation of an observation. In the logic-based approach to abduction, the background theory is given by a consistent set of formulas Σ. The notion of a *possible explanation* is defined by saying that a formula γ is an explanation of α if $\Sigma \cup \{\gamma\} \vdash \alpha$. An explanatory relation is a binary relation $\rhd$ where the intended meaning of $\alpha \rhd \gamma$ is "γ is a *preferred explanation* of α".

Structural properties for abduction has been studied by Flach [3], Cialdea-Mayer and Pirri [10], Aliseda [1]. The search for these properties is motivated by questions of the following type: (i) Suppose that γ is a preferred explanation of $\alpha \wedge \beta$. Should γ be considered also a preferred explanation of α? (ii) If γ is a preferred explanation of α and also of β, is γ a preferred explanation of $\alpha \vee \beta$? (iii) If γ is a preferred explanation of α and γ' entails γ, should γ' be considered a preferred explanation of α? Answers to these questions will tell how much a change of an observation affects its preferred explanations and more important will contribute to truly make abduction a form of logical inference. We have presented in [13, 14] a fairly complete list of rationality postulates for abduction in the Kraus-Lehmann-Magidor tradition.

A second natural question is whether changes in the background theory should be allowed during the explanatory process. We will not consider this interesting aspect in this paper (some remarks about it can be found in the introduction of [14]).

In this paper we will consider a third aspect of abduction: preference criteria for selecting explanations. Most formalism have treated them as external devices which work on top of the logical part of abduction. However, the exact relationship between the preference criteria and the logical or structural properties of the explanatory mechanism has not been so far clearly delineated. The main goal of this paper is to clarify this problem.

Perhaps the most natural way of defining an explanatory relation $\rhd$ is through a preference relation $\prec$ over formulas. The relation $\prec$ will tell which formulas in $Expla(\alpha)$ (the set of possible explanations of α) are the preferred ones. Let us define $\alpha \rhd \gamma$ iff γ is a $\prec$-minimal element of $Expla(\alpha)$. We will show the exact correspondence between general structural properties for $\rhd$ and the properties of the preference relation $\prec$ (i.e, properties like being modular, filtered, smooth, etc). In particular, we will show that the preference criteria is, in fact, implicit in the structural rules satisfied by a given explanatory mechanism. Moreover, we will show that this is necessarily the case: the logical properties satisfied by an explanatory relation $\rhd$ already encode a selection mechanism. These are the main results of this paper.

This paper is organized as follows: In Section 2 we recall the main structural rules introduced in [14] and the basic hierarchy of explanatory relations. Section 3 deals with explanatory relations defined by "orders" (selection mechanisms) over formulas. We define the essential relation and prove some basic representation theorems. Section 4 is devoted to study the role of the particular rule *Right And.* In Section 5 we present some examples. In Section 6 we compare briefly our work with other related works. We conclude with some remarks and open problems in Section 7.

2 Structural rules

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

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

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

Definition 2.1. *For every formula α, the collection of explanations of α w.r.t. Σ is denoted by $Expla(\alpha)$ and is defined as follows:*

$$Expla(\alpha) = \{\gamma : \gamma \nvdash_\Sigma \bot \quad \& \quad \gamma \vdash_\Sigma \alpha\}$$

Notice that we have ruled out trivial explanations by asking that γ has to be consistent with Σ. We are interested in studying the relation "γ is a preferred explanation of α", which will be denoted by $\alpha \rhd \gamma$. In explanatory reasoning the input is an observation and the output is an explanation, that is the reason to write $\alpha \rhd \gamma$ with α as input and γ as output. Our next definition capture the ideas mentioned in the introduction.

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

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

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

The following rules were introduced in [14] and are the structural rules mentioned in the introduction. These rules are desirable since in a sense they describe properties of well behaved explanatory relations:

LLE$_\Sigma$: $$\frac{\vdash_\Sigma \alpha \leftrightarrow \alpha' \ ; \ \alpha \rhd \gamma}{\alpha' \rhd \gamma}$$

RLE$_\Sigma$: $$\frac{\vdash_\Sigma \gamma \leftrightarrow \gamma' \ ; \ \alpha \rhd \gamma}{\alpha \rhd \gamma'}$$

E-CM: $$\frac{\alpha \rhd \gamma \ ; \ \gamma \vdash_\Sigma \beta}{(\alpha \wedge \beta) \rhd \gamma}$$

E-C-Cut: $$\frac{(\alpha \wedge \beta) \rhd \gamma \ ; \ \forall \delta \, [\alpha \rhd \delta \ \Rightarrow \ \delta \vdash_\Sigma \beta\,]}{\alpha \rhd \gamma}$$

RA: $$\frac{\alpha \rhd \gamma \ ; \ \gamma' \vdash_\Sigma \gamma \ ; \ \gamma' \nvdash_\Sigma \bot}{\alpha \rhd \gamma'}$$

LOR: $$\frac{\alpha \rhd \gamma \ ; \ \beta \rhd \gamma}{(\alpha \vee \beta) \rhd \gamma}$$

E-DR: $$\frac{\alpha \rhd \gamma \ ; \ \beta \rhd \delta}{(\alpha \vee \beta) \rhd \gamma \text{ or } (\alpha \vee \beta) \rhd \delta}$$

E-R-Cut: $$\frac{(\alpha \wedge \beta) \rhd \gamma \ ; \ \exists \delta \, [\alpha \rhd \delta \ \& \ \delta \vdash_\Sigma \beta]}{\alpha \rhd \gamma}$$

E-RW : $$\frac{\alpha \rhd \gamma \ ; \ \alpha \rhd \delta}{\alpha \rhd (\gamma \vee \delta)}$$

E-Con$_\Sigma$: $\not\vdash_\Sigma \neg\alpha$ iff there is γ such that $\alpha \rhd \gamma$

E-Reflexivity: $$\frac{\alpha \rhd \gamma}{\gamma \rhd \gamma}$$

Notice that some of these postulates (like E-C-Cut) are not rules in the usual finitary sense, since they have quantifiers in the premises. However, for the sake of simplicity, we will keep the standard notation used for rules in propositional logic.

The justification and intuition behind these rules were given in [14]. Nevertheless, for the sake of completeness, we will make some brief comments about these rules. The rules (LLE$_\Sigma$) *Left Logical Equivalence* and (RLE$_\Sigma$) *Right Logical Equivalence* are very natural assumptions. They say that explanatory relations are independent of the syntax. The rule (E-CM), *Explanatory Cautious Monotony*, expresses a weak form of a monotonicity on the left.

The rules (E-C-Cut), *Explanatory Cautious Cut*, and (E-R-Cut), *Explanatory Rational Cut* are the explanatory cut rules. They play an important role in our setting. Actually, there is a duality between monotony rules for consequence relations and cut rules. They say that a preferred explanation of the more complex observation $(\alpha \wedge \beta)$ might also be, in some cases, a preferred explanation of the simpler or incomplete observation (α). In other words, Cut rules allow to keep a preferred explanation even when the set of observations is not longer complete. As we will see, these rules reflect a selection mechanism. In fact, a failure of full cut (*i.e.* $\alpha \wedge \beta \rhd \gamma$ but $\alpha \not\rhd \gamma$) says that there is some part (β) of an observation $(\alpha \wedge \beta)$ which is so relevant for explaining the whole observation that it can not be ignored. This difference between the whole observation and a part of it will be reflected in the selection of preferred explanations.

The rule RA, *Right And*, gives some amount of monotony on the right. A similar postulate has been considered by Flach in [3]. RA says that any explanation more "complete"(logically stronger, more specific) than a preferred explanation of α is also a preferred explanation of α. Notice that RA and E-RW are the only rules that introduce new explanations.

Let us remark some controversial aspects of RA. A typical example where one might not wish to have RA is something of the following sort: if *"there is sugar in the coffee"* is a preferred explanation of the observation *"the coffee is good"* then RA would declare that *"there are sugar and pepper in the coffee"* should also be a preferred explanation of the observation. It is in cases like this that the theory Σ can play an important role by ruling out such undesirable explanations. However, it is clear that RA fails when every observation has an unique (up to logical equivalence) preferred explanation. There are natural examples of such explanatory relations. In Section 4 we will study more deeply the rule RA.

The rule (LOR) is *Left Or*. The intuition behind this rule is the following. Suppose that when we observe either α or β (no matter which one) we are willing to accept that γ is a very likely explanation for both of them. Now we are told that one of them is observed (but maybe it is not known which one) then LOR says that it is rational to conclude that γ is still a very likely explanation of that observation (*i.e.* a very likely explanation of $\alpha \vee \beta$). The rule E-DR *Explanatory Disjunctive Rationality* is stronger than LOR and has a similar interpretation.

Finally E-Con$_\Sigma$, *Explanatory Consistency Preservation*, is the postulate that says when a formula has a preferred explanation: just when the observation is consistent with Σ.

Definition 2.3. *Let Σ be a background theory and $\rhd$ be an explanatory relation. We say that $\rhd$ is* **E-preferential** *if it satisfies* E-CM, E-C-Cut, LLE$_\Sigma$ *and* RA. $\rhd$ *is* **E-disjunctive rational** *if it is E-preferential and in addition satisfies* E-DR. $\rhd$ *is* **E-rational** *if it is E-preferential and in addition satisfies* E-R-Cut.

The basic motivation is the following. To each explanatory relation $\rhd$ we associate a consequence relation $\mid\!\sim_{ab}$ as follows:

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

We read $\alpha \mid\!\sim_{ab} \beta$ as "normally, when α is observed then β should also be present".

We proved in [14] that if $\rhd$ is E-preferential then $\mid\!\sim_{ab}$ is preferential (in the KLM sense [6]); if $\rhd$ is E-disjunctive rational then $\mid\!\sim_{ab}$ is disjunctive rational (*i.e.* in addition of the preferential rules the following also holds: if $\alpha \vee \beta \mid\!\sim \rho$ then $\alpha \mid\!\sim \rho$ or $\beta \mid\!\sim \rho$) and finally if $\rhd$ is E-rational then $\mid\!\sim_{ab}$ is rational (*i.e.* a preferential relation which satisfies also *rational monotony* :if $\alpha \mid\!\sim \rho$ and $\alpha \not\mid\!\sim \neg\rho$ then $\alpha \wedge \beta \mid\!\sim \rho$). These notions form a hierarchy:

E-preferential $\subset$ E-disjunctive rational $\subset$ E-rational.

To each consequence relation $\mid\!\sim$ we associate an explanatory relation $\tilde{\rhd}$ by putting

$$\alpha \tilde{\rhd} \gamma \quad \text{iff } \gamma \not\vdash_{\Sigma} \bot \text{ and } C(\alpha) \subseteq Cn\ (\Sigma \cup \{\gamma\}) \tag{2}$$

For an adequate consequence relation $\mid\!\sim$ [1] we have shown [14] that if $\mid\!\sim$ is preferential then $\tilde{\rhd}$ is E-preferential; if $\mid\!\sim$ is disjunctive rational then $\tilde{\rhd}$ is E-disjunctive rational and if $\mid\!\sim$ is rational then $\tilde{\rhd}$ is E-rational.

These results show a formal duality between explanatory relations and consequence relations. We will give more details later in the paper.

3 Ordering explanations.

As we have said in the introduction the most distinct feature of abduction is the emphasis it makes on preferred explanations rather than plain explanations. In this section we will focus on preference criteria for defining explanatory relations. We will show that these preference criteria are implicitly built in the structural properties of explanatory relations introduced in Section 2. These results are quite natural on the light of the well known facts about non-monotonic reasoning. In fact, it is well known that inference processes based on orders over formulas are one the "faces" of non monotonic reasoning [11]. For instance possibilistic orders [2] and expectations orders [5] characterize inference rational relations. Preferential orders [4] characterize preferential relations. We will comment about their connection with our results.

We will start by making precise some basic notions. If $\prec$ is an irreflexive binary relation over a set S and $A \subseteq S$, then $a \in A$ 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 used we will just write $min(A)$.

We formally define the notion of a preference relation.

[1] $\mid\!\sim$ is said to be ***adequate with respect to*** Σ if for every formula α the following holds:

$$C(\alpha) = \bigcap \{Cn\ (\Sigma \cup \{\gamma\}) :\ C(\alpha) \subseteq Cn(\Sigma \cup \{\gamma\})\}$$

Definition 3.1. *A* **preference relation** $\prec$ *will be any binary irreflexive relation* $\prec$ *over* $\mathcal{L}$ *which is invariant under logical equivalence w.r.t.* Σ, i.e. *if* $\alpha \prec \beta$ *and* $\vdash_\Sigma \alpha \leftrightarrow \alpha'$ *and* $\vdash_\Sigma \beta \leftrightarrow \beta'$, *then* $\alpha' \prec \beta'$.

To each preference relation $\prec$ we associate an explanatory relation $\rhd$ as follows:

Definition 3.2. *Let* $\prec$ *be an irreflexive relation on formulas. The explanatory relation* $\rhd$ *associated to* $\prec$ *is defined by*

$$\alpha \rhd \gamma \overset{def}{\Leftrightarrow} \gamma \in min(Expla(\alpha), \prec) \tag{3}$$

i.e. $\alpha \rhd \gamma$ *iff* $\not\vdash_\Sigma \neg\gamma$, $\gamma \vdash_\Sigma \alpha$ *and* $\delta \not\vdash_\Sigma \alpha$ *for all* δ *such that* $\delta \prec \gamma$.

Definition 3.2 is the same one given in [8, 10] (but notice that they worked with reflexive relations). Notice also that $\prec$ is not supposed to be transitive, thus the notion of $\prec$-minimal element might not be intuitive. We will be interested mainly in the case where $\prec$ is at least smooth (see definition below). The two basic questions that we will address are then the following:

1. To determine the relationship between the structural properties of $\prec$ (like being a partial, filtered, modular order) and the structural rules satisfied by $\rhd$.

2. To determine under which conditions a given explanatory relation can be represented by a preference relation $\prec$ as in (3).

First, we point out a simple fact.

Proposition 3.3. *Let* $\prec$ *be a binary relation as in 3.2 and* $\rhd$ *be the corresponding explanatory relation. Then* $\rhd$ *satisfies* E-Reflexivity *and* E-CM.

Proof: To see **E-Reflexivity** just notice that if $\alpha \rhd \gamma$ then it is obvious that $\gamma \in min(Expla(\gamma)), \prec)$.To check that E-CM holds, suppose $\alpha \rhd \gamma$ and $\gamma \vdash_\Sigma \beta$, then $\gamma \in min(Expla(\alpha)) \cap Expla(\alpha \wedge \beta) \subseteq min(Expla(\alpha \wedge \beta))$. ∎

To obtain other postulates we will impose some constrains over $\prec$. The following notion of a smooth relation is inspired by the notion of smoothness used in the study of consequence relations ([6]).

Definition 3.4. *Let* $\prec$ *be a reflexive 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* $\prec$ *as in 3.1 is called* **smooth**, *if for every formula* α *the set* $Expla(\alpha)$ *is smooth.*

To understand better the meaning of smoothness we remark the following: Let $A \subseteq B \subseteq S$, then it is clear that $min(B) \cap A \subseteq min(A)$. 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, $min(A) \subseteq min(B)$. Notice that when the language if finite every transitive relation $\prec$ is obviously smooth.

Theorem 3.5. *If $\prec$ is a smooth preference relation and $\rhd$ is defined as in (3.2), then $\rhd$ is an explanatory relation that satisfies* LLE_Σ, RLE_Σ, E-CM, E-C-Cut *and* $\mathsf{E\text{-}Con}_\Sigma$.

Proof: That LLE_Σ and RLE_Σ hold follows from the fact that $\prec$ is logically invariant. We already have shown in 3.3 that E-CM holds. To see that $\mathsf{E\text{-}Con}_\Sigma$ holds, suppose that α is consistent with Σ then $Expla(\alpha)$ is not empty. By smoothness there is γ such that $\alpha \rhd \gamma$. To see that E-C-Cut, suppose that the premises in the rule E-C-Cut hold. Hence $min(Expla(\alpha)) \subseteq Expla(\beta)$ and since $Expla(\alpha \wedge \beta) \subseteq Expla(\alpha)$, then $min(Expla(\alpha)) \subseteq min(Expla(\alpha \wedge \beta))$. Since $Expla(\alpha)$ is smooth we conclude $min(Expla(\alpha)) = min(Expla(\alpha \wedge \beta))$ and this finishes the proof. ∎

It seems natural to expect that under the conditions in the conclusion of 3.5 the relation $\rhd$ is represented by a smooth preference relation as in 3.2. However, in order to get such representation we will need more than just E-CM and E-C-Cut.

First, we introduce some necessary notions.

Definition 3.6. *Let $\rhd$ be an explanatory relation. We will say that a formula γ is* admissible for $\rhd$ *if $\alpha \rhd \gamma$ for some formula α.*

The following definition, motivated by the results in [12], is the key point in order to get our basic representation theorem.

Definition 3.7. *Let $\rhd$ be an explanatory relation that satisfies* RLE_Σ. *The* essential preference relation *associated to $\rhd$ is denoted by $\prec_e$ and defined by:*

(a) For δ not admissible: $\gamma \prec_e \delta$ for every admissible γ.

(b) For γ and δ admissible: $\gamma \prec_e \delta$ if $Cn(\Sigma \cup \{\gamma\}) \cap \{\beta : \beta \rhd \delta\} = \emptyset$.

The only relevant formulas for the definition of $\prec_e$ are admissible formulas. Since $\rhd$ satisfies RLE_Σ then $\prec_e$ is invariant under logically equivalence and thus it is indeed a preference relation. Notice that admissible formulas are consistent with Σ. Admissible formulas play in our paper the same role as normal models in [6, 9]. A concept similar to that of an admissible formula was defined in [3].

Remark 3.8. *Suppose $\prec$ is a preference relation and $\rhd$ is the associated explanatory relation (3.2). It is easy to verify that if $\gamma \prec \delta$, then $\gamma \prec_e \delta$. In other words, $\prec_e$ is larger than $\prec$.*

The proof that $\prec_e$ represents $\rhd$ will work when the language is finite and more generally for explanatory relations which are logically finite either on the right or on the left (see definition below). First, we introduce an auxiliary notion.

Definition 3.9. *A set of formulas A is said to have an upper bound (in A w.r.t Σ) if there are finitely many formulas $\alpha_1, \ldots, \alpha_n \in A$ such that for all $\alpha \in A$, $\alpha \vdash_\Sigma (\alpha_1 \vee \cdots \vee \alpha_n)$ (i.e, $\alpha_1 \vee \cdots \vee \alpha_n$ is an upper bound of A in the lattice of formulas modulo Σ).*

Definition 3.10. *An explanatory relation $\rhd$ is said to be* **logically finite on the right** *(RLF) if for every formula α the set $\{\gamma : \alpha \rhd \gamma\}$ has an upper bound.*

Definition 3.11. *An explanatory relation $\rhd$ is said to be* **logically finite on the left** *(LLF) if for every admissible formula γ the set $\{\alpha : \alpha \rhd \gamma\}$ has an upper bound.*

Definition 3.12. *An explanatory relation $\rhd$ is said to be* **logically finite** *if it satisfies RLF or LLF.*

Notice that if the language is finite then every explanatory relation is logically finite. We will give two examples of logically finite relations:

Example 3.13. *Let $\mid\sim$ be an adequate consequence relation and $\tilde{\rhd}$ be the explanatory relation associated to $\mid\sim$, defined by (2). If $\tilde{\rhd}$ is logically finite on the right, then there is a map F from formulas into formulas such that $C(\alpha) = Cn(\Sigma \cup \{F(\alpha)\})$. In fact, for every α let $F(\alpha)$ be $\gamma_1 \vee \cdots \vee \gamma_n$ the upper bound for $\{\gamma : \alpha \rhd \gamma\}$ given by 3.10 (if α is inconsistent with Σ, then we let $F(\alpha)$ be $\bot$). Conversely, it is clear that if such function F exists then $\tilde{\rhd}$ satisfies RLF.*

Example 3.14. *We will present an example of a LLF explanatory relation $\tilde{\rhd}$. We define first an adequate rational relation $\mid\sim$ as follows: Consider an infinite language $\mathcal{L} = \{p_1, p_2, \ldots\}$. Let $\mathcal{L}_n = \{p_1, p_2, \ldots, p_n\}$ and fix m models $M_1, \ldots, M_m$ for the language $\mathcal{L}_n$. Let $L'_1 \cdots, L'_k$ be a partition of $\{M_1, \ldots, M_m\}$ in k levels and let $L'_i = \{M_i^1, \ldots, M_i^{n_i}\}$ for $i = 1, \ldots, k$. Now consider the ranked model in the language $\mathcal{L}$ given by k levels $L_1, \ldots, L_k$, where $M \in L_i$ iff the restriction of M to $\mathcal{L}_n$ is in L'_i. Let γ_i^j be formulas in* *the language $\mathcal{L}_n$ such that* $\mathrm{mod}(\gamma_i^j) = M_i^j$. *For* $r = 1, \ldots, k$ *we let* β_r *be the following formula:*

$$\beta_r = \bigvee_{i=r}^{k} (\bigvee_{j=1}^{n_i} \gamma_i^j)$$

Let $\Sigma = \{\beta_1\}$. *It is not hard to see that the rational relation* $\mid\!\sim$ *generated by this model is adequate (with respect to* Σ*). Moreover, this ranked model is standard,* i.e. *for every formula* α, $\mathrm{mod}(C(\alpha)) = \mathrm{mod}(\alpha) \cap L_i$, *where* i *is the first integer* j *such that* $\mathrm{mod}(\alpha) \cap L_j \neq \emptyset$. *It is easy to check that a formula* γ *is admissible iff* $C(\gamma) = Cn(\Sigma \cup \{\gamma\})$. *Let* γ *be an admissible formula, then there is* r *such that* $\mathrm{mod}(\Sigma \cup \{\gamma\}) \subseteq L_r$. *We claim that* β_r *is an upper bound for* $\{\alpha : \alpha \tilde{\rhd} \gamma\}$. *In fact, it is easy to see that* $\mathrm{mod}(\beta_r) = \bigcup_{i=r}^{k} L_i$ *and* $\mathrm{mod}(C(\beta_r)) = L_r$. *Hence* $\beta_r \tilde{\rhd} \gamma$. *Now, if* $\alpha \tilde{\rhd} \gamma$, *then* $\mathrm{mod}(\Sigma \cup \{\alpha\}) \subseteq \bigcup_{i=r}^{k} L_i$. *Thus* $\alpha \vdash_\Sigma \beta_r$.

The next theorem gives a characterization of those logically finite explanatory relations representable by preference relations.

Theorem 3.15. *Let* $\rhd$ *be a logically finite explanatory relation. The following are equivalent:*

(i) The relation $\rhd$ *satisfies* E-CM, LLE_Σ, RLE_Σ, E-C-Cut, $\mathsf{E\text{-}Con}_\Sigma$ *and* LOR.

(ii) There is a smooth preference relation $\prec$ *such that*

$$min(Expla(\alpha)) \cap min(Expla(\beta)) \subseteq min(Expla(\alpha \vee \beta)) \quad (4)$$

and for every formula α *the following holds*

$$\alpha \rhd \gamma \quad \textit{iff} \quad \gamma \in min(Expla(\alpha), \prec) \quad (5)$$

Proof: $(ii) \Rightarrow (i)$. By 3.5 we only need to show that $\rhd$ satisfies LOR. But this follows immediately from (4).

$(i) \Rightarrow (ii)$. We will show that $\prec_e$ works. We already have observed that since $\rhd$ satisfies RLE_Σ then $\prec_e$ is a preference relation. First, notice that (4) follows immediately from (5) and LOR.

We will show that (5) holds. Let us suppose that $\alpha \rhd \gamma$ and let $\delta \in Expla(\alpha)$, then $\alpha \in Cn(\Sigma \cup \{\delta\}) \cap \{\beta : \beta \rhd \gamma\}$. Therefore $\delta \not\prec_e \gamma$ and $\gamma \in min(Expla(\alpha))$. This shows that the *only if* in (5) holds.

Fix a formula α' consistent with Σ and let δ' be any formula in $Expla(\alpha')$. We will show that if $\alpha' \not\triangleright \delta'$, then there is γ 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 (5) holds. Suppose $\alpha' \not\triangleright \delta'$. If δ' is not admissible, then there is nothing to show because of the definition of $\prec_e$ and **E-Con**$_\Sigma$. Hence we will assume that δ' is admissible. By **E-Con**$_\Sigma$ there is γ such that $\alpha' \triangleright \gamma$, so let

$$C_{\alpha'} = \bigcap \{Cn\ (\Sigma \cup \{\gamma\}) : \alpha' \triangleright \gamma\}$$

and

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

We claim that S is consistent. In fact, suppose towards a contradiction, that S is inconsistent. By compactness there are β_i's for $i = 1, \ldots, n$ such that $\beta_i \triangleright \delta'$ and $(\beta_1 \vee \cdots \vee \beta_n) \in C'_\alpha$. 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.

Since $\triangleright$ is logically finite there are two cases to be considered:

(a) $\triangleright$ satisfies **RLF**, *i.e.* for every formula α the set $A = \{\gamma : \alpha \triangleright \gamma\}$ has an upper bound. Let $\gamma_i \in A$, $i \leq n$ be an upper bound for A. It is easy to check that

$$\begin{aligned} C_{\alpha'} &= \bigcap \{Cn\ (\Sigma \cup \{\gamma\}) : \gamma \in A\} \\ &= \bigcap \{Cn\ (\Sigma \cup \{\gamma_i\}) : i \leq n\} \\ &= Cn\ (\Sigma \cup \{(\gamma_1 \vee \cdots \vee \gamma_n)\}. \end{aligned}$$

Let N be a model of S, then there is i such that $N \models \Sigma \cup \{\gamma_i\}$. As N is also a model of $\{\neg\beta : \beta \triangleright \delta'\}$, then it is clear that $\gamma_i \prec_e \delta'$.

(b) $\triangleright$ satisfies **LLF**, *i.e.* for every admissible formula γ the set $\{\beta : \beta \triangleright \gamma\}$ has a upper bound. Since δ' is admissible, 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 β such that $\beta \triangleright \delta'$. Let $\beta' = \beta_1 \vee \cdots \vee \beta_n$, then by **LOR** $\neg\beta' \in S$. Since S is consistent then $\beta' \notin C_{\alpha'}$, hence there is γ such that $\alpha' \triangleright \gamma$ and $\gamma \not\vdash_\Sigma \beta'$. Therefore $\gamma \not\vdash_\Sigma \beta$, for all β such that $\beta \triangleright \delta'$, *i.e.* $\gamma \prec_e \delta'$. ∎

We will continue now analyzing the properties that $\prec_e$ has when $\triangleright$ satisfies extra axioms. We postpone to Section 4 the analysis of the effect that **RA** has on $\prec_e$.

When $\triangleright$ satisfies **E-DR**, then $\prec_e$ can be described in a different way (a similar idea was used in [9, 12]). Recall that from Section 2 we know that **E-DR** implies **LOR**. We introduce the following definition

Definition 3.16. *Let* $\rhd$ *be an explanatory relation that satisfies* RLE_Σ*. Define a binary relation* $\prec_u$ *by:*

(a) For δ *not admissible:* $\gamma \prec_u \delta$ *for every admissible* γ*.*

(b) For γ *and* δ *admissible:*

$$\gamma \prec_u \delta \stackrel{def}{\Leftrightarrow} \forall\alpha\forall\beta\,[\alpha \rhd \gamma \;\&\; \beta \rhd \delta \;\Rightarrow\; (\alpha \vee \beta) \rhd \gamma \;\&\; (\alpha \vee \beta) \not\rhd \delta]$$

Proposition 3.17. *Let* $\rhd$ *be an explanatory relation that satisfies* LLE_Σ*,* RLE_Σ*, E-CM, E-C-Cut, and E-DR. Then* $\prec_e = \prec_u$*. Moreover,* $\prec_u$ *(and therefore* $\prec_e$*) is transitive.*

Proof: ($\prec_e \subseteq \prec_u$): This follows quite straightforward from the hypotheses.

($\prec_u \subseteq \prec_e$): Let γ, δ be admissible formulas with $\gamma \prec_u \delta$. Suppose, towards a contradiction, that there is β such that $\beta \rhd \delta$ and $\gamma \vdash_\Sigma \beta$. Let α be any formula such that $\alpha \rhd \gamma$. Since $\gamma \vdash_\Sigma \beta$, then by E-CM we have $(\alpha \wedge \beta) \rhd \gamma$. Since $\vdash ((\alpha \wedge \beta) \vee \beta) \leftrightarrow \beta$ and $\gamma \prec_u \delta$, then (by LLE_Σ) we conclude that $\beta \not\rhd \delta$, which is a contradiction.

To see that $\prec_u$ is transitive, let γ_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 γ_i is admissible. Let α_i be formulas such that $\alpha_i \rhd \gamma_i$. By E-DR it suffices to show that $(\alpha_1 \vee \alpha_3) \not\rhd \gamma_3$. Suppose, towards a contradiction, that $(\alpha_1 \vee \alpha_3) \rhd \gamma_3$. Since $\gamma_2 \prec_u \gamma_3$, then by definition of $\prec_u$ we have $(\alpha_1 \vee \alpha_2 \vee \alpha_3) \rhd \gamma_2$ and $(\alpha_1 \vee \alpha_2 \vee \alpha_3) \not\rhd \gamma_3$. Since $\gamma_1 \prec_u \gamma_2$, then analogously we have $(\alpha_1 \vee \alpha_2 \vee \alpha_3) \rhd \gamma_1$ and $(\alpha_1 \vee \alpha_2 \vee \alpha_3) \not\rhd \gamma_2$, which is a contradiction. ∎

In [4] it was used a notion of filtered relation. We can adapt this notion to our context as follows:

Definition 3.18. *A preference relation* $\prec$ *is said to be* filtered *if for every* α *and every* $\gamma, \gamma' \in Expla(\alpha)$ *such that* $\gamma \notin min(Expla(\alpha))$ *and* $\gamma' \notin min(Expla(\alpha))$*, there is* $\delta \in min(Expla(\alpha))$*, such that* $\delta \prec \gamma$ *and* $\delta \prec \gamma'$*.*

Using an argument similar to that in the proof of 3.15 the following theorem can be proved:

Theorem 3.19. *If* $\rhd$ *is a logically finite explanatory relation that satisfies* RLE_Σ*, E-CM, E-C-Cut,* $\mathsf{E\text{-}Con}_\Sigma$*, E-RW and E-DR then* $\prec_u$ *(therefore* $\prec_e$*) is filtered.*

4 The role of RA

In general, the information contained in an explanatory relation $\rhd$ could be lost when we move to its associate consequence relation $\mid\!\sim_{ab}$ (defined in (1)).

However, as it was shown in [14], this is not the case for *causal* explanatory relations which are those satisfying the following condition:

$$\alpha \rhd \gamma \text{ iff } C_{ab}(\alpha) \subseteq Cn(\Sigma \cup \gamma) \tag{6}$$

Where $C_{ab}(\alpha) = \{\beta : \alpha \mid\!\sim_{ab} \beta\}$. In other words, $\rhd$ is causal if $\rhd = \tilde{\rhd}$ where $\tilde{\rhd}$ is the explanatory relation associated to $\mid\!\sim_{ab}$ which justifies our claim that for causal explanatory relations $\mid\!\sim_{ab}$ and $\rhd$ contains the same information. For a finite language, we have shown that an explanatory relation is causal iff it satisfies RA and E-RW. It is then clear that RA must have a very distinct effect on $\prec_e$. Causal explanations relations are nonmonotonic reasoning in reverse [14]

The following proposition says that if RA holds, then $\prec_e$ satisfies almost all the properties of a *preferential pre-ordering* as defined by Freund in [4]. In Section 6 we will compare in more detail the properties of preferential orders and the essential relation $\prec_e$.

Proposition 4.1. *Let* $\rhd$ *be an explanatory relation that satisfies* RLE_Σ. *Then the following holds:*

(i) Let γ, γ' and δ be admissible formulas such that $\gamma \vdash \gamma'$. If $\gamma \prec_e \delta$, then $\gamma' \prec_e \delta$.

(ii) Let γ and δ be admissible formulas. If $(\delta \vee \gamma) \prec_e \gamma$, then $\delta \prec_e \gamma$.

(iii) Suppose that $\rhd$ *also satisfies* RA. *Let γ, γ' and δ be formulas such that $\gamma \not\vdash_\Sigma \bot$ and $\gamma \vdash \gamma'$. If $\delta \prec_e \gamma$, then $\delta \prec_e \gamma'$.*

Proof: To see (ii), suppose $\delta \not\prec_e \gamma$ and let β be such that $\delta \vdash_\Sigma \beta$ and $\beta \rhd \gamma$. Then clearly $\gamma \vee \delta \vdash_\Sigma \beta$ and thus $(\gamma \vee \delta) \not\prec_e \gamma$. The proof of (i) is similar. For (iii), suppose that $\delta \prec_e \gamma$. Thus by definition δ is admissible. If γ' is not admissible then by definition $\delta \prec_e \gamma'$. Now suppose that γ' is admissible and also, towards a contradiction, that $\delta \not\prec_e \gamma'$. Let β be such that $\beta \rhd \gamma'$ and $\delta \vdash_\Sigma \beta$. Since $\gamma \vdash_\Sigma \gamma'$ and $\gamma \not\vdash_\Sigma \bot$, we have by RA that $\beta \rhd \gamma$ and therefore $\delta \not\prec_e \gamma$. ∎

Remark: *A way of understanding part (iii) of the previous proposition is as follows: Assume that* $\rhd$ *satisfies* RLE_Σ *and* RA. *Let γ_1 and γ_2 be two admissible formulas such that $\gamma_1 \vee \gamma_2$ is also admissible. Let $\gamma' = \gamma_1 \vee \gamma_2$. It is easy to check using 4.1 that $\gamma_i \not\prec_e \gamma'$ and $\gamma' \not\prec_e \gamma_i$,* i.e. *for each i, γ_i and γ' are $\prec_e$-incomparable. But in fact, 4.1 (i) (resp. (iii)) says more, namely that every formula above (resp. below) γ_i is also above (resp. below) $\gamma_1 \vee \gamma_2$. So in some sense $\gamma_1 \vee \gamma_2$ contains the information "coded" by γ_1 and γ_2.*

Since explanatory relations are defined using $\prec$-minimal explanations it is clear the relevance of (iii).

Property (iii) corresponds to RA and we will denote this property by C-U (Continuing Up). To define it formally, we say that a binary relation $<$ over formulas satisfies C-U if the following holds:

C-U $\gamma \not\vdash_\Sigma \bot \;\&\; \gamma \vdash \gamma' \;\&\; \delta < \gamma \;\Rightarrow \delta < \gamma'$

Proposition 4.2. *If $\prec$ is a preference relation satisfying* C-U *then the explanatory relation associated to $\prec$ (defined in 3.2) satisfies* RA.

Proof: Suppose that $\alpha \rhd \gamma$, $\gamma' \vdash_\Sigma \gamma$ and $\gamma' \not\vdash_\Sigma \bot$. We want to show that $\alpha \rhd \gamma'$, *i.e.* $\gamma' \in min(Expla(\alpha), \prec)$. Since $\gamma' \not\vdash_\Sigma \bot$ then it is clear that $\gamma' \in Expla(\alpha)$. For reductio, assume there is $\delta \in Expla(\alpha)$ such that $\delta \prec \gamma'$. By C-U and since $\gamma' \vdash_\Sigma \gamma$ we have $\delta \prec \gamma$ contradicting the minimality of γ in $Expla(\alpha)$. ∎

In the result that follows, it is interesting to notice that the hypothesis of logically finiteness is not needed. We will use this result in the sequel.

Proposition 4.3. *Let $\rhd$ be an E-rational explanatory relation satisfying* E-Con$_\Sigma$. *Then the following holds: for all admissible formulas γ and δ,*

$$\gamma \prec_u \delta \Leftrightarrow \exists\alpha\,\exists\beta\;[\,\alpha \rhd \gamma \;\&\; \beta \rhd \delta \;\&\; (\alpha\vee\beta) \rhd \gamma \;\&\; (\alpha\vee\beta) \not\rhd \delta\,] \tag{7}$$

Moreover, $\prec_u$ (and therefore $\prec_e$) is smooth and represents $\rhd$.

Proof: The $\Rightarrow$ direction comes directly from the definition of $\prec_u$. For the other direction, let α and β be as in the right hand side of (7) and α' and β' be formulas such that $\alpha' \rhd \gamma$ and $\beta' \rhd \delta$. We need to show that $(\alpha' \vee \beta') \rhd \gamma$ and $(\alpha' \vee \beta') \not\rhd \delta$. Since $\rhd$ is E-rational, $\rhd$ satisfies E-DR (see [14]), hence it suffices to show that $(\alpha' \vee \beta') \not\rhd \delta$. Suppose, towards a contradiction, that $(\alpha' \vee \beta') \rhd \delta$. By E-CM we have $(\alpha' \vee \beta') \wedge (\alpha \vee \beta) \rhd \delta$. And by hypothesis $(\alpha \vee \beta) \rhd \gamma$ and clearly $\gamma \vdash_\Sigma (\alpha' \vee \beta')$, hence by E-R-Cut $(\alpha \vee \beta) \rhd \delta$, which contradicts the choice of α and β.

To see that $\prec_u$ is smooth, we first recall that $\rhd$ satisfies E-DR and therefore, by 3.17, $\prec_u = \prec_e$. As in the proof of 3.15 we have that if $\alpha \rhd \gamma$ then $\gamma \in min(Expla(\alpha), \prec_e)$. For the other direction, let $\delta \in Expla(\alpha)$ such that $\alpha \not\rhd \delta$. We will find γ such that $\alpha \rhd \gamma$ and $\gamma \prec_u \delta$. This will show that $\prec_u$ is smooth and also that it represents $\rhd$. We can assume without lost of generality that δ is admissible and thus let β be such that $\beta \rhd \delta$. Hence by

E-CM $(\alpha \wedge \beta) \rhd \delta$. By E-Con$_\Sigma$ there is γ such that $\alpha \rhd \gamma$. Since α is logically equivalent to $(\alpha \wedge \beta) \vee \alpha$, then $((\alpha \wedge \beta) \vee \alpha) \not\rhd \delta$ and $((\alpha \wedge \beta) \vee \alpha) \rhd \gamma$. From (7) we conclude that $\gamma \prec_u \delta$. This finishes the proof. ∎

We will show next that when $\rhd$ satisfies E-R-Cut then $\prec_u$ is *modular.* We recall the definition of a modular relation (see [7]):

Definition 4.4. *A relation $\prec$ on a set E is said to be* modular *iff there exists a linear order $<$ on some set Ω and a function $r : E \to \Omega$ such that $a \prec b$ iff $r(a) < r(b)$. If $\prec$ is transitive, modularity is equivalent to the following property: for all a, b and c in E if a and b are $\prec$-incomparable and $a \prec c$ then $b \prec c$.*

Theorem 4.5. *Let $\rhd$ be an explanatory relation, the following are equivalent:*

(i) The relation $\rhd$ is E-rational and satisfies E-Con$_\Sigma$.

(ii) There is a smooth and modular preference relation $\prec$ satisfying C-U *such that for every α we have*

$$\alpha \rhd \gamma \quad \textit{iff} \quad \gamma \in min(Expla(\alpha), \prec)$$

Proof: (i $\Rightarrow$ ii) From 3.15 we know that $\prec_e$ represent $\rhd$. From 4.1 we know that C-U holds. Thus, it remains to see that $\prec_e$ (alias $\prec_u$) is modular. Let γ, δ and ρ be formulas such that $\gamma \not\prec_u \delta$, $\delta \not\prec_u \gamma$ and $\gamma \prec_u \rho$. We want to show that $\delta \prec_u \rho$. Without lost of generality we can assume that γ, δ and ρ are admissible. Let α, β, ω formulas such that $\alpha \rhd \gamma$, $\beta \rhd \delta$ and $\omega \rhd \rho$. Since γ and δ are $\prec_u$-incomparable then from 4.3 it follows that $(\alpha \vee \beta) \rhd \gamma$ and $(\alpha \vee \beta) \rhd \delta$. Again by 4.3 it suffices to show that $(\alpha \vee \beta \vee \omega) \rhd \delta$ and $(\alpha \vee \beta \vee \omega) \not\rhd \rho$. By E-DR, which is true because $\rhd$ is E-rational, it is enough to show that $(\alpha \vee \beta \vee \omega) \not\rhd \rho$. Since $\gamma \prec_u \rho$, then by definition of $\prec_u$ we have $(\alpha \vee \beta \vee \omega) \rhd \gamma$ and $(\alpha \vee \beta \vee \omega) \not\rhd \rho$.

(ii $\Rightarrow$ i) From 3.5 we know that $\rhd$ satisfies LLE$_\Sigma$, RLE$_\Sigma$, E-CM, E-C-Cut and E-Con$_\Sigma$. From 4.2 we obtain RA. It remains to be shown that E-R-Cut holds. Let α, β, γ and δ formulas such that $(\alpha \wedge \beta) \rhd \gamma$, $\alpha \rhd \delta$ and $\delta \vdash_\Sigma \beta$. We need to show that $\alpha \rhd \gamma$. Suppose, towards a contradiction, that $\alpha \not\rhd \gamma$. Since $\gamma \vdash_\Sigma \alpha$, then by smoothness and the definition of $\rhd$, there is δ' such that $\alpha \rhd \delta'$ and $\delta' \prec \gamma$. Since $\alpha \rhd \delta$ then $\delta \not\prec \delta'$ and $\delta' \not\prec \delta$. By E-CM $(\alpha \wedge \beta) \rhd \delta$ and by modularity, $\delta \prec \gamma$, which contradicts the hypothesis that $(\alpha \wedge \beta) \rhd \gamma$. ∎

5 Examples

Preferential models are perhaps the easiest way of defining structures for modelling various knowledge representation problems. We can also use them here to construct explanatory relations illustrating the properties of $\prec_e$.

We will work with a finite language. A *preferential model* [2] consists of a collection S of valuations and a partial order $\prec$ on S. In our case S will be the collection $\text{mod}(\Sigma)$ of models of Σ. Given a formula α we define its minimal models as usual:

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

The relation $\prec$ over valuations is meant to capture the preferences of the agent and thus $\min(\alpha)$ contains the most preferred or normal worlds where the observation α holds. Therefore we can use $\min(\alpha)$ to capture also our preference over explanations. There are several ways of doing so. We will present three of them.

In order to get a more clear picture of the examples that follow it is also convenient to have in mind the consequence relation associated to the preferential model which is defined by

$$\alpha \mid\sim \beta \text{ iff } \min(\alpha) \subseteq \text{mod}(\beta) \tag{8}$$

The following definition is also useful

$$C(\alpha) = \{\beta : \alpha \mid\sim \beta\}$$

$C(\alpha)$ contains the nonmonotonic consequences of α.

(1) Causal explanatory relations: Define an explanatory relation $\rhd_c$ as follows:

$$\alpha \rhd_c \gamma \overset{def}{\Leftrightarrow} mod(\Sigma \cup \{\gamma\}) \subseteq min(\alpha) \tag{9}$$

for any pair of consistent (with Σ) formulas α and γ. In other words, an explanation of α is a preferred one if all its models are normal for α. Notice that $\alpha \rhd_c \gamma$ iff $C(\alpha) \subseteq Cn(\Sigma \cup \gamma)$ and $\mid\sim_{ab} = \mid\sim$. Hence we have that $\rhd_c$ is equal to $\tilde{\rhd}$ for the consequence relation $\mid\sim$ given in (8) and therefore $\rhd_c$ is a causal relation (as defined in (6)).

It is not difficult to show that this method always yields E-preferential explanatory relations that moreover satisfies LOR and E-RW. Since the

[2] For the more general definition of a preferential model see [6]

hypothesis of theorem 3.15 holds then $\rhd_c$ is represented by its associated essential relation which will be denoted by $\prec_e^c$. We will make some remarks about this particular $\prec_e^c$.

Admissible formulas are given by the antichains of the preferential model (*i.e.* sets of mutually $\prec$-incompatible valuations). In other word, γ is $\rhd_c$-admissible iff $mod(\Sigma \cup \{\gamma)\}$ is an antichain. And conversely, given a collection $\mathcal{A} \subseteq \text{mod}(\Sigma)$ of mutually incompatible valuations, then any formula γ such that $\mathcal{A} = \text{mod}(\gamma)$ is $\rhd_c$-admissible. For $\rhd_c$-admissible formulas γ and δ the following holds:

$$\gamma \prec_e^c \delta \text{ iff } \exists N \models \delta \, \exists M \models \gamma \text{ such that } M \prec N$$

Notice that, in general, the relation $\prec_e^c$ is clearly not transitive. In fact it is quite easy to find an example such that $\gamma \prec_e^c \delta$ and conversely $\delta \prec_e^c \gamma$. However, the representation theorem 3.15 guarantees that $\prec_e^c$ is smooth. In the case that the preferential model is filtered (for the definition see [4]) then $\prec_e^c$ is transitive (this follows from 3.17 and the fact that in this case **E-DR** holds).

(2) Strong epistemic explanatory relations: Consider now the following explanatory relation

$$\alpha \rhd_{SE} \gamma \overset{def}{\Leftrightarrow} \gamma \vdash_\Sigma \alpha \;\;\&\;\; \min(\gamma) \subseteq \min(\alpha) \tag{10}$$

In words, an explanation of α is a preferred one if all its minimal models are also minimal for α Notice that $\alpha \rhd_{SE} \gamma$ iff $C(\alpha) \subseteq C(\gamma)$ and $\gamma \vdash_\Sigma \alpha$. More details and motivations about this notion are given in §4 of [14]. For instance, $\rhd_{SE}$ satisfies **LLE**, **RLE**, **E-CM**, **E-RW**, **E-C-Cut** but **RA** does not hold. Notice that $\rhd_{SE}$ is full reflexive, so every formula consistent with Σ is $\rhd_{SE}$-admissible.

The relation $\prec_e^s$ is characterized in a way quite similar to that of $\prec_e^c$.

$$\gamma \prec_e^s \delta \text{ iff } \exists N \in \min(\delta) \, \exists M \models \gamma \text{ such that } M \prec N$$

So the crucial difference is the notion of an admissible formula. In general **LOR** might not hold for $\rhd_{SE}$ (so theorem 3.15 does not apply) however $\rhd_{SE}$ is represented by its associated essential relation $\prec_e^s$ *i.e.* $\alpha \rhd_{SE} \gamma$ iff γ is a $\prec_e^s$-minimal explanation of α.

(3) NMC explanatory relations: Consider the following:

$$\alpha \rhd_{nc} \gamma \overset{def}{\Leftrightarrow} \gamma \vdash_\Sigma \alpha \;\&\; \text{mod}(\gamma) \cap \min(\alpha) \neq \emptyset \tag{11}$$

for any pair of consistent (with Σ) formulas α and γ. In words, an explanation of α is a preferred one if at least one of its models is minimal for α. Notice that this is equivalent to saying that $\gamma \vdash_\Sigma \alpha$ and $\alpha \not\mid\!\sim \neg\gamma$, so we called $\rhd_{nc}$ *nonmonotonically consistent* explanatory relation. It is not difficult to show that $\rhd_{nc}$ satisfies LLE, RLE,E-CM, E-RW, E-C-Cut but RA does not hold. Notice that $\rhd_{nc}$ is full reflexive, so every formula consistent with Σ is $\rhd_{nc}$-admissible.

The essential relation $\prec_e^{nc}$ is characterized as follows.

$$\gamma \prec_e^{nc} \delta \text{ iff } \forall N \in \min(\delta)\ \exists M \in \min(\gamma) \text{ with } M \prec N$$

Notice that $\prec_e^{nc}$ is transitive. Similar to what happens with the strong epistemic relation, $\rhd_{nc}$ might not satisfies LOR (and hence theorem 3.15 does not apply) however it is representable by $\prec_e^{nc}$. Moreover, $\prec_e^{nc}$ is a well known order among formulas as we will see in the following section.

6 Related works

We will comment briefly in this section about the connection of our results and the works of Freund [4], Gärdenfors and Makinson [5] and Dubois and Prade [2].

Freund characterize preferential consequence relations in terms of 'preferential orders'. He called *preferential order* any relation $<$ on formulae satisfying the following four properties:

P_0: $\alpha < \bot$
P_1: If $\alpha \vdash \beta$, then (a) $\alpha < \gamma \Rightarrow \beta < \gamma$
(b) $\delta < \beta \Rightarrow \delta < \alpha$
P_2: If $\alpha < \gamma$ and $\alpha < \delta$, then $\alpha < \gamma \vee \delta$
P_3: If $\alpha \vee \beta < \beta$, then $\alpha < \beta$

The connection of preferential orders with preferential consequence relations is as follows: Given a preferential consequence relation $\mid\!\sim$ define $<_{\mid\sim}$ by letting $\alpha <_{\mid\sim} \beta$ if $\alpha \vee \beta \mid\!\sim \neg\beta$. And conversely, given a preferential order $<$ on formulae define a consequence relation $\mid\!\sim_<$ by letting $\alpha \mid\!\sim_< \beta$ if $\alpha < \alpha \wedge \neg\beta$. Freund showed that $<_{\mid\sim}$ is a preferential order and $\mid\!\sim_<$ is a preferential consequence relation. The connection of Freund's order with our work is the following: for the NMC explanatory relation defined in Section 5 we have that $\prec_e^{nc}$ is exactly Freund's relation.

In general, the essential relation $\prec_e$ associated to a given explanatory relation $\rhd$ satisfies P_0 when the formulas α and β are admissible and, except for $\mathsf{P}_{\mathsf{l}}(\mathrm{b})$, the others properties are also satisfied by $\prec_e$ (this follows from 4.1). However, it is important to remark that in general $\prec_e$ is not transitive but Freund's relation is. The conditions **C-U** and $\mathsf{P}_{\mathsf{l}}(\mathrm{b})$ seem to play dual roles, however a complete classification of preference relations is still to be done.

Finally, the expectation orders of Gärdenfors and Makinson are modular and can be defined in terms of Freund's relation as follows: $<$ is an expectation order iff the dual relation $<^*$ is a modular preferential order, where $\alpha <^* \beta$ if $\neg\beta < \neg\alpha$. A similar result holds for the possibilistic order of Dubois and Prade.

7 Final remarks

Selection mechanisms are a fundamental part of abduction. However, most formalism have treated them as external devices which work on top of the logical part of abduction. We have shown that preference criteria are built in the structural properties of explanatory relations. Moreover, our results show that the preference criteria has to be somewhat uniform in order that an explanatory relation satisfies structural rules.

There are some natural questions suggested by our results. First of all, our representation theorem 3.15 is not optimal, since we have presented examples of explanatory relations which are representable by its associated essential relation but they do not satisfy **LOR**. Secondly, all examples we have examined so far are based on preferential models. It would be interesting, for a future work, to study preference relations defined in terms of a *simplicity* criteria, for instance, syntactic simplicity. Finally, it would be also interesting to find representation theorems capturing exactly the type of explanatory relations defined by (10) and (11).

Acknowledgment: This work has been partially supported by a CNRS (France)- CONICIT (Venezuela) grant # PI98003494.

References

[1] Atocha Aliseda-Llera. *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence.* PhD thesis, Stanford University, Department of Computer Science, 1998.

[2] D. Dubois and H. Prade. Epistemic entrenchment and possibilistic logic. *Artificial Intelligence*, 50:223–239, 1991.

[3] P. A. Flach. Rationality postulates for induction. In Yoav Shoham, editor, *Proc. of the Sixth Conference of Theoretical Aspects of Rationality and Knowledge (TARK96)*, pages 267–281, The Netherlands, March 17-20, 1996.

[4] M. Freund. Injective models and disjunctive relations. *Journal of Logic and Computation*, 3:231–247, 1993.

[5] P. Gärdenfors and D. Makinson. Nonmonotonic inferences based on expectations. *Artificial Intelligence*, 65:197–245, 1994.

[6] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[7] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55:1–60, 1992.

[8] H.J. Levesque. A knowledge level account of abduction. In *Proceedings of the eleventh International Joint Conference on Artificial Intelligence*, pages 1061–1067, Detroit, 1989.

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

[10] M. Cialdea Mayer and F. Pirri. Abduction is not deduction-in-reverse. *Journal of the IGPL*, 4:1–14, 1996.

[11] D. Makinson. Five faces of minimality. *Studia Logica*, 52:339–379, 1993.

[12] R. Pino Pérez and C. Uzcátegui. On representation theorems for nonmonotonic inference relations. *Journal of Symbolic Logic*. To appear.

[13] R. Pino Pérez and C. Uzcátegui. Abduction vs. deduction in nonmonotonic reasoning. In J. Delgrande and M. Truszczynski, editors, *Proceedings of the Seventh International Workshop on Nonmonotonic Reasoning, NR'98*, pages 42–54, Trento, Italy, May 30-June 1, 1998.

[14] R. Pino-Pérez and C. Uzcátegui. Jumping to explanations vs. jumping to conclusions. *Artificial Intelligence*, 111(2):131–169, 1999.

Ramón Pino-Pérez Visiting the Departamento de Matemáticas, Facultad de Ciencias, Universidad de Los Andes, Mérida 5101, Venezuela (until August 2000)

Centre de recherches en informatique de Lens
Faculté des Sciences
Université d'Artois
62307 Lens, France
pino@cril.univ-artois.fr
and
Carlos Uzcátegui
Departamento de Matemáticas
Facultad de Ciencias
Universidad de Los Andes
Mérida 5101, Venezuela
uzca@ciens.ula.ve