International Journal of Approximate Reasoning 102 (2018) 1–20

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Explanatory relations in arbitrary logics based on satisfaction systems, cutting and retraction ☆

Marc Aiguier [a], Jamal Atif [b], Isabelle Bloch [c,*], Ramón Pino Pérez [d]

[a] MICS, Centrale Supelec, Université Paris-Saclay, France
[b] Université Paris-Dauphine, PSL Research University, CNRS, UMR 7243, LAMSADE, 75016 Paris, France
[c] LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France
[d] Departemento de Matemáticas, Facultad de Ciencias, Universidad de Los Andes, Mérida 5101, Venezuela

ARTICLE INFO

Article history:
Received 28 February 2018
Received in revised form 6 June 2018
Accepted 31 July 2018
Available online 6 August 2018

Keywords:
Explanatory relation
Retraction
Cutting
Satisfaction systems

ABSTRACT

The aim of this paper is to introduce a new framework for defining abductive reasoning operators based on a notion of retraction in arbitrary logics defined as satisfaction systems. We show how this framework leads to the design of explanatory relations satisfying properties of abductive reasoning, and discuss its application to several logics. This extends previous work on propositional logics where retraction was defined as a morphological erosion. Here weaker properties are required for retraction, leading to a larger set of suitable operators for abduction for different logics.

1. Introduction

Since its introduction by Charles Peirce in [34], abduction has motivated a large body of research in several scientific fields, e.g. philosophy of science, logics, law, artificial intelligence, to mention a few. Abduction, whatever the adopted view on its treatment, involves a background theory (T), an observation also called explanandum (φ), and an explanation (ψ). The observation may be seen as a surprising phenomenon that is inconsistent with the background theory. It may also be consistent with the background theory but not directly entailed by this theory, which is the case considered in this paper. Several constraints can be imposed on the explanations and on the process of their production. One can allow changing the background theory, or not, consider as non relevant explanations those that entail the observation on their own without engaging the background knowledge. Hence, several forms of abduction can be defined depending on the chosen criteria. Despite their divergence, most of these models agree to define abduction as an explanatory reasoning allowing us to infer the best explanations of an observation. This contributes to the field of explainable artificial intelligence. Explanatory relations, trying to model common sense and everyday reasoning, find applications in many domains, such as diagnosis [16,22], forensics [30], argumentation [11,12], language understanding [31], image understanding [4,41], etc. (it is out of the scope of this paper to describe applications exhaustively). Then, as a form of inference, several rationality postulates have been studied, that are more appropriate to govern the process of selecting the best explanations, e.g. [24,

☆ This paper is part of the Virtual special issue on Defeasible and Ampliative Reasoning, Edited by Ivan Varzinczak, Richard Booth, Giovanni Casini, Szymon Klarman and Gilles Richard.

* Corresponding author.
E-mail addresses: marc.aiguier@centralesupelec.fr (M. Aiguier), jamal.atif@dauphine.fr (J. Atif), isabelle.bloch@telecom-paristech.fr (I. Bloch), pino@ula.ve (R. Pino Pérez).

https://doi.org/10.1016/j.ijar.2018.07.014
0888-613X/

35]. From a computational point of view, a very large number of papers has tackled the definition of abductive procedures, mainly in propositional logics. An attractive approach, governed by what is called the AKM model, is based on semantic tableaux tailored for particular logics (e.g. propositional logics [3], first order and modal logics [14,15]), which was the basis for several extensions (e.g. [5,22,28,37]). In these approaches, the explanatory reasoning process is split into two stages: (i) generating a set of hypotheses from the formulas that allow closing the open branches in the tableau constructed from $(T \cup \{\neg\varphi\})$, and (ii) selecting the preferred solutions from this plain set by considering some of the criteria mentioned above.

Our aim in this paper is to introduce a new framework for defining abductive reasoning operators in arbitrary logics in the framework of satisfaction systems. To this end, we propose a new notion of cutting, from which operators of retraction are derived. We show that this framework leads to the design of explanatory relations satisfying the rationality postulates of abductive reasoning introduced in [35] and adapted here to the proposed more general framework, and present applications in several logics. This extends previous work on abduction in propositional logics where retraction was defined as a morphological erosion [8–10], as well as abduction in description logics for image understanding [4]. Here weaker properties are required for retraction, that allow defining a larger set of suitable operators for abduction for different logics. This approach is similar to the one proposed for revision in [1], where revision operators were defined from relaxation in satisfaction systems, and then instantiated in various logics. An important feature of the proposed explanations based on retraction is that generation and selection steps are merged, or at least the set of generated hypotheses is reduced, thus facilitating the selection step.

The paper is organized as follows. In Section 2, we recall the useful definitions and properties of satisfaction systems, and provide examples in propositional logic, Horn logic, first order logic, modal propositional logic and description logic. In Section 3 we introduce our first contribution, by defining a notion of cutting, from which explanations are then defined. In Section 4, we propose to define particular cuttings, based on retractions of formulas. Then in Section 5, we instantiate the proposed general framework in various logics.

2. Satisfaction systems

We recall here the basic notions of satisfaction systems needed in this paper. The presentation follows the one in [1], where we give a more complete presentation of satisfaction systems, including the properties and their proofs, that are omitted here.

2.1. Definition and examples

Definition 1 *(Satisfaction system).* A **satisfaction system** $\mathcal{R} = (Sen, Mod, \models)$ consists of

- a set Sen of **sentences**,
- a class Mod of **models**, and
- a satisfaction relation $\models \subseteq Mod \times Sen$.

Let us note that the non-logical vocabulary, so-called *signature*, over which sentences and models are built, is not specified in Definition 1.[1] Actually, it is left implicit. Hence, as we will see in the examples developed in the paper, a satisfaction system always depends on a signature.

Example 1. The following examples of satisfaction systems are of particular importance in computer science and in the remainder of this paper.

Propositional Logic (PL) Given a set of propositional variables Σ, we can define the satisfaction system $\mathcal{R}_\Sigma = (Sen, Mod, \models)$ where Sen is the least set of sentences finitely built over propositional variables in Σ, the symbols $\top$ and $\bot$ (denoting tautologies and antilogies – or contradictions –, respectively), and Boolean connectives in $\{\neg, \vee, \wedge, \Rightarrow\}$, Mod contains all the mappings $\nu : \Sigma \to \{0, 1\}$ (0 and 1 are the usual truth values), and the satisfaction relation $\models$ is the usual propositional satisfaction.

Horn Logic (HCL) A *Horn clause* is a sentence of the form $\Gamma \Rightarrow \alpha$ where Γ is a finite (possibly empty) conjunction of propositional variables and α is a propositional variable. The satisfaction system of Horn clause logic is then defined as for **PL** except that sentences are restricted to be conjunctions of Horn clauses.

Modal Propositional Logic (MPL) Given a set of propositional variables Σ, we can define the satisfaction system $\mathcal{R}_\Sigma = (Sen, Mod, \models)$ where

- Sen is the least set of sentences finitely built over propositional variables in Σ, the symbols $\top$ and $\bot$, Boolean connectives in $\{\neg, \vee, \wedge, \Rightarrow\}$, and modalities in $\{\Box, \Diamond\}$;
- Mod contains all the Kripke models (I, W, R) where I is an index set, $W = (W^i)_{i \in I}$ is a family of functions from Σ to $\{0, 1\}$, and $R \subseteq I \times I$ is an accessibility relation;

[1] The set of logical symbols is defined in each particular logic and does not depend on a theory.

- the satisfaction of sentences by Kripke models, $(I, W, R) \models \varphi$, is defined by $(I, W, R) \models_i \varphi$ for every $i \in I$ where $\models_i$ is defined by structural induction on sentences as follows:
  - $(I, W, R) \models_i p$ iff $p \in W^i$ for every $p \in \Sigma$,
  - Boolean connectives are handled as usual,
  - $(I, W, R) \models_i \Box\varphi$ iff $(I, W, R) \models_j \varphi$ for every $j \in I$ such that $(i, j) \in R$, and
  - $\Diamond\varphi$ is the same as $\neg\Box\neg\varphi$.

First Order Logic (FOL) and Many-sorted First Order Logic We detail here only the many-sorted variant of FOL, FOL being a particular case. Signatures are triplets (S, F, P) where S is a set of sorts, and F and P are sets of function and predicate names, respectively, each with arity in $S^* \times S$ and S^+ respectively (S^+ is the set of all non-empty sequences of elements in S and $S^* = S^+ \cup \{\epsilon\}$ where ϵ denotes the empty sequence). In the following, to indicate that a function name $f \in F$ (respectively a predicate name $p \in P$) has for arity $(s_1 \dots s_n, s)$ (respectively $s_1 \dots s_n$), we will write $f : s_1 \times \dots \times s_n \to s$ (respectively $p : s_1 \times \dots \times s_n$).

Given a signature $\Sigma = (S, F, P)$, we can define the satisfaction system $\mathcal{R}_\Sigma = (Sen, Mod, \models)$ where:

- Sen is the least set of sentences built over atoms of the form $p(t_1, \dots, t_n)$ where $p : s_1 \times \dots \times s_n \in P$ and $t_i \in T_F(X)_{s_i}$ for every i, $1 \le i \le n$ ($T_F(X)_s$ is the term algebra of sort s built over F with sorted variables in a given set X) by finitely applying Boolean connectives in $\{\neg, \vee, \wedge, \Rightarrow\}$ and quantifiers in $\{\forall, \exists\}$.
- Mod is the class of models $\mathcal{M}$ defined by a family $(M_s)_{s \in S}$ of non-empty sets (one for every $s \in S$), each one equipped with a function $f^{\mathcal{M}} : M_{s_1} \times \dots \times M_{s_n} \to M_s$ for every $f : s_1 \times \dots \times s_n \to s \in F$ and with an n-ary relation $p^{\mathcal{M}} \subseteq M_{s_1} \times \dots \times M_{s_n}$ for every $p : s_1 \times \dots \times s_n \in P$.
- Finally, the satisfaction relation $\models$ is the usual first-order satisfaction.

As for **PL**, we can consider the logic **FHCL** of first-order Horn Logic whose models are those of **FOL** and sentences are restricted to be conjunctions of universally quantified Horn sentences (i.e. sentences of the form $\Gamma \Rightarrow \alpha$ where Γ is a finite conjunction of atoms and α is an atom).

Description logic (DL) Signatures are triplets (N_C, N_R, I) where N_C, N_R and I are nonempty pairwise disjoint sets where elements in N_C, N_R and I are called concept names, role names and individuals, respectively.

Given a signature $\Sigma = (N_C, N_R, I)$, we can define the satisfaction system $\mathcal{R}_\Sigma = (Sen, Mod, \models)$ where:

- Sen contains[2] all the sentences of the form $C \sqsubseteq D$, $x : C$ and $(x, y) : r$ where $x, y \in I$, $r \in N_R$ and C is a concept inductively defined from $N_C \cup \{\top\}$ and binary and unary operators in $\{\_ \sqcap \_, \_ \sqcup \_\}$ and in $\{\neg\_, \forall r.\_, \exists r.\_\}$, respectively.
- Mod is the class of models $\mathcal{I}$ defined by a set $\Delta^{\mathcal{I}}$ equipped for every concept name $A \in N_C$ with a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, for every relation name $r \in N_R$ with a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and for every individual $x \in I$ with a value $x^{\mathcal{I}} \in \Delta^{\mathcal{I}}$.
- The satisfaction relation $\models$ is then defined as:
  - $\mathcal{I} \models C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$,
  - $\mathcal{I} \models x : C$ iff $x^{\mathcal{I}} \in C^{\mathcal{I}}$,
  - $\mathcal{I} \models (x, y) : r$ iff $(x^{\mathcal{I}}, y^{\mathcal{I}}) \in r^{\mathcal{I}}$,

  where $C^{\mathcal{I}}$ is the evaluation of C in $\mathcal{I}$ inductively defined on the structure of C as follows:
  - if $C = A$ with $A \in N_C$, then $C^{\mathcal{I}} = A^{\mathcal{I}}$;
  - if $C = \top$ then $C^{\mathcal{I}} = \Delta^{\mathcal{I}}$;
  - if $C = C' \sqcup D'$ (resp. $C = C' \sqcap D'$), then $C^{\mathcal{I}} = C'^{\mathcal{I}} \cup D'^{\mathcal{I}}$ (resp. $C^{\mathcal{I}} = C'^{\mathcal{I}} \cap D'^{\mathcal{I}}$);
  - if $C = \neg C'$, then $C^{\mathcal{I}} = \Delta^{\mathcal{I}} \setminus C'^{\mathcal{I}}$;
  - if $C = \forall r.C'$, then $C^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \forall y \in \Delta^{\mathcal{I}}, (x, y) \in r^{\mathcal{I}} \text{ implies } y \in C'^{\mathcal{I}}\}$;
  - if $C = \exists r.C'$, then $C^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}}, (x, y) \in r^{\mathcal{I}} \text{ and } y \in C'^{\mathcal{I}}\}$.

2.2. Knowledge bases and theories

Let us now consider a fixed but arbitrary satisfaction system $\mathcal{R} = (Sen, Mod, \models)$ (since the signature Σ is supposed fixed, the subscript Σ will be omitted from now on).

Notation 1. Let $T \subseteq Sen$ be a set of sentences.

- $Mod(T)$ is the sub-class of Mod whose elements are models of T, i.e. $Mod(T) = \{\mathcal{M} \in Mod \mid \forall \varphi \in T, \mathcal{M} \models \varphi\}$. When T is restricted to a formula φ (i.e. $T = \{\varphi\}$), we will denote the class of models of $\{\varphi\}$ by $Mod(\varphi)$, rather than $Mod(\{\varphi\})$.
- $Cn(T) = \{\varphi \in Sen \mid \forall \mathcal{M} \in Mod(T),\ \mathcal{M} \models \varphi\}$ is the set of *semantic consequences of* T. In the following, we will also denote $T \models \varphi$ to mean that $\varphi \in Cn(T)$.
- $\varphi \equiv_T \psi$ iff $Mod(T \cup \{\varphi\}) = Mod(T \cup \{\psi\})$.
- Let $\mathbb{M} \subseteq Mod$. Let us define $\mathbb{M}^* = \{\varphi \in Sen \mid \forall \mathcal{M} \in \mathbb{M}, \mathcal{M} \models \varphi\}$. When $\mathbb{M}$ is restricted to one model $\mathcal{M}$, $\mathbb{M}^*$ will be equivalently noted $\mathcal{M}^*$.

[2] The description logic defined here is better known under the acronym $\mathcal{ALC}$.

- Let us note $Triv = \{\mathcal{M} \in Mod \mid \mathcal{M}^* = Sen\}$, i.e. the set of models in which all formulas are satisfied. In **PL**, **MPL** and **FOL**, $Triv$ is empty because the negation is considered. Similarly, the negation is involved in the **DL** $\mathcal{ALC}$, hence $Triv$ is empty. In **HCL**, $Triv$ only contains the unique model where all propositional variables have a truth value equal to 1. In **FHCL**, $Triv$ contains all models $\mathcal{M}$ where for every predicate name $p : s_1 \times \ldots \times s_n \in P$, $p^{\mathcal{M}} = M_{s_1} \times \ldots \times M_{s_n}$.

Definition 2 *(Knowledge base and theory).* A **knowledge base (KB)** T is a finite set of sentences (i.e. $T \subseteq Sen$ and the cardinality of T belongs to $\mathbb{N}$). A set of sentences T is said to be a **theory** if and only if $T = Cn(T)$.

A theory T is **finitely representable** if there exists a KB $T' \subseteq Sen$ such that $T = Cn(T')$.

A class of models $\mathbb{M} \subseteq Mod$ is **finitely axiomatizable** if there exists a finite KB T such that $Mod(T) = \mathbb{M}$. A satisfaction system $\mathcal{R}$ is **finitely axiomatizable** if each of its classes of models $\mathbb{M} \subseteq Mod$ is finitely axiomatizable.

Note that in DL, a knowledge base consists classically of a set of axioms (of the form $C \sqsubseteq D$), called TBox, and a set of assertions (of the form $x : C$ or $(x, y) : r$), called ABox.

Classically, the consistency of a theory T is defined as $Mod(T) \neq \emptyset$. The problem of such a definition of consistency is that its significance depends on the considered logic. Hence, this consistency is significant for **FOL**, while in **FHCL** it is a trivial property since each set of sentences is consistent because $Mod(T)$ always contains $Triv$ which is non empty. Here, for the notion of consistency to be more appropriate for our purpose of defining abduction for the largest family of logics, we propose a more general definition of consistency, the meaning of which is that given a theory T, $Mod(T)$ is not restricted to trivial models.

Definition 3 *(Consistency).* $T \subseteq Sen$ is **consistent** if $Cn(T) \neq Sen$.

Accordingly, inconsistency of T means $Cn(T) = Sen$.

Proposition 1 *([1]). For every $T \subseteq Sen$, T is consistent if and only if $Mod(T) \setminus Triv \neq \emptyset$.*

Note that this definition of consistency is equivalent to $Triv \subset Mod(T)$, with $\subset$ denoting the strict inclusion. Hence, for every $T \subseteq Sen$, T is inconsistent is equivalent to $Mod(T) = Triv$.

2.3. Internal logic

Following [17,27], the satisfaction system-independent definition of Boolean connectives is straightforward. This will be useful when we give general results of preserving explanatory relations along Boolean connectives. Let $\mathcal{R}$ be a satisfaction system. A sentence φ' is a

- **semantic negation** of φ when $Mod(\varphi') = Mod \setminus Mod(\varphi)$;
- **semantic conjunction** of φ_1 and φ_2 when $Mod(\varphi') = Mod(\varphi_1) \cap Mod(\varphi_2)$;
- **semantic disjunction** of φ_1 and φ_2 when $Mod(\varphi') = Mod(\varphi_1) \cup Mod(\varphi_2)$;
- **semantic implication** of φ_1 and φ_2 when $Mod(\varphi') = (Mod \setminus Mod(\varphi_1)) \cup Mod(\varphi_2)$.

The satisfaction system $\mathcal{R}$ has (semantic) negation when each sentence has a negation. It has (semantic) conjunction (respectively disjunction and implication) when any two sentences have conjunction (respectively disjunction and implication). As usual, we note negation, conjunction, disjunction and implication by $\neg$, $\wedge$, $\vee$ and $\Rightarrow$.

Example 2. **PL** has all semantic Boolean connectives. **FOL** has all semantic Boolean connectives when sentences are restricted to closed formulas, otherwise (i.e. sentences can be open formulas) it only has semantic conjunction. Finally, **MPL** has only semantic conjunction.

3. Explanation in satisfaction systems

The process of inferring the best explanation of an observation is usually known as *abduction*. In a logic-based approach, the background of abduction is given by a knowledge base (KB) T and a formula φ (the observation) such that $T \cup \{\varphi\}$ is consistent. Besides this fact, which can be expressed equivalently as $T \not\models \neg\varphi$, it is sometimes further required that $T \not\models \varphi$. We do not impose this last requirement here.

Let us start by introducing the notion of explanation of φ with respect to T.

Definition 4 *(Set of explanations).* Let T be a KB. Let $\varphi \in Sen$ be a formula consistent with T. The **set of explanations** of φ over T is the set $Expla_T(\varphi)$ defined as:

$$Expla_T(\varphi) = \{\psi \in Sen \mid Mod(T \cup \{\psi\}) \neq Triv \text{ and } T \cup \{\psi\} \models \varphi\}$$

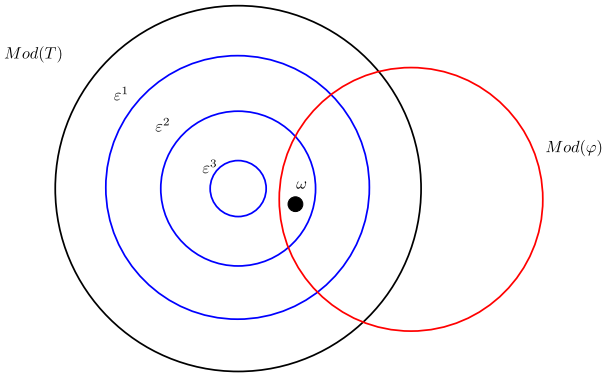


Fig. 1. Schematic illustration of the idea of explanations derived from morphological erosions: the set of models $Mod(T)$ is iteratively eroded, providing an ordered sequence of subsets of models $\varepsilon^3 \subseteq \varepsilon^2 \subseteq \varepsilon^1 \subseteq Mod(T)$. A possible way to find explanations of φ is to identify the most central models in $Mod(T)$ which are also models of φ. Here ω is such a model. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Note that this definition does not impose that $\psi \not\models \varphi$. In some cases a preferred explanation of φ with respect to the background knowledge base T could be a formula ψ such that $\psi \models \varphi$.

Since abduction aims to infer the best explanations, the notion of explanation given in Definition 4 only captures candidate explanations of φ with respect to T. Some additional properties are needed to define the key notion of "preferred explanations". Following the work in [3,24–26,35,36], we will study some preference criteria and give their logical properties when abduction is regarded as a form of inference.

Definition 5 *(Explanatory relation).* Let T be a KB. An **explanatory relation for** T is a binary relation $\rhd \subseteq Sen \times Sen$ such that:

$$\forall \varphi, \psi \in Sen, \varphi \rhd \psi \Longrightarrow \psi \in Expla_T(\varphi)$$

Now, we define an (abstract) explanatory relation, the behavior of which will consist in cutting in $Mod(T \cup \{\varphi\})$ as much as possible but still under the constraint that it remains consistent (i.e. it is not equal to $Triv$). A cutting will then generate a sequence of subsets of $\mathcal{P}(Mod(T \cup \{\varphi\}))$ that we can order by inclusion. Moreover, this sequence cannot be extended by inverse inclusion.

This idea comes from our previous work on morphological erosions of propositional formulas to express the most central models, representing robust explanations [9]. The sequence of erosions of increasing size provides an ordering on the set of models. This is illustrated in Fig. 1.

Here we generalize this idea and propose more abstract notions, that can be instantiated in various logics. The notion of cutting introduced in the next definition comes from the idea of erosion, which, in some way, cuts in the set of models. This gives rise to the notion of a cutting for a KB T and a formula φ. Examples of cutting based on the notion of retraction (close to the one of morphological erosion, with slightly different properties) are provided in Section 4, some instantiations in various logics are described in Section 5.

Definition 6 *(Cutting).* Let T be a KB and let φ be a formula. A **cutting** for T and φ is any $\mathcal{C} \subseteq \mathcal{P}(Mod(T \cup \{\varphi\}))$ such that for every $\mathbb{M} \in \mathcal{C}$, $Triv \subset \mathbb{M}$ (and then $\mathbb{M} \setminus Triv \neq \emptyset$), $\mathcal{C}$ is closed under set-theoretical union and contains $Mod(T \cup \{\varphi\})$, and the poset $(\mathcal{C}, \subseteq)$ is well-founded.[3]

Let us denote by $Min(\mathcal{C})$ the set of minimal elements for $\subseteq$ in $\mathcal{C}$.

In the following, given a KB T and a formula φ, a cutting for T and φ will be denoted by $\mathcal{C}_\varphi$.[4]

Note that in Definition 6, we do not impose that $T \cup \{\varphi\}$ is consistent but this is implicit since the definition implies that $Mod(T \cup \{\varphi\}) \setminus Triv \neq \emptyset$. The case where $T \cup \{\varphi\}$ is not consistent is not interesting and not considered for explanatory relations.

The fact that a cutting is closed under union and well-founded will be useful next to define maximal chains with minimal elements, in a way similar to the successive erosions in Fig. 1. An example of not well-founded poset could be $(\mathcal{P}(Mod), \subseteq)$ for an infinite set Mod (which is hence not a cutting).

[3] Let us recall that a poset $(X, \preceq)$ is well-founded if every non-empty subset $S \subseteq X$ has a minimal element with respect to $\preceq$, or equivalently there does not exist any infinite descending chain.

[4] To simplify the notations, T does not index cuttings because as we will see, T will be often constant.

Remark 1. If $Mod(T \cup \{\varphi\}) \neq Triv$ then there exists a trivial cutting for φ, namely $\mathcal{C}_\varphi = \{Mod(T \cup \{\varphi\})\}$, which means that the whole set of models of $T \cup \{\varphi\}$ is kept, i.e. none is "cut".

As $(\mathcal{C}_\varphi, \subseteq)$ is closed under set-theoretical union and therefore it is inductive, by the Hausdorff maximal principle, every chain is contained in some maximal chain (and therefore maximal chains exist). Moreover, as $(\mathcal{C}_\varphi, \subseteq)$ is well-founded, every maximal chain has a least element which belongs to $Min(\mathcal{C}_\varphi)$.

Definition 7 *(Explanatory relation based on cuttings).* Let T be a KB, and let us define a set of cuttings $\mathcal{C}$ by choosing a cutting $\mathcal{C}_\varphi$ for every φ in Sen: $\mathcal{C} = \{\mathcal{C}_\varphi \mid \varphi \in Sen\}$. Let us define the binary relation $\rhd_\mathcal{C} \subseteq Sen \times Sen$ as follows:

$$\varphi \rhd_\mathcal{C} \psi \iff \begin{cases} Mod(T \cup \{\psi\}) \setminus Triv \neq \emptyset, \text{ and} \\ \exists \mathbb{M} \in Min(\mathcal{C}_\varphi), Mod(T \cup \{\psi\}) \subseteq \mathbb{M} \end{cases}$$

This definition makes sense since, by Remark 1, there is at least one $\mathcal{C}_\varphi$ for each φ which is consistent. Obviously, $\rhd_\mathcal{C}$ is an explanatory relation. We will later add some stability properties to $\mathcal{C}$ to ensure good properties of this explanatory relation.

Remark 2. If $\mathcal{C}_\varphi$ is a cutting for T and φ, then we can define a relation $\rhd_\mathcal{C}$ based on cuttings such that $\varphi \rhd_\mathcal{C} \psi$ satisfies the equivalence of Definition 7 (i.e. $\mathcal{C}_\varphi$ is precisely the cutting chosen for φ in the set $\mathcal{C}$).

Note that the explanatory relation depends on the minimal elements of a cutting. The other elements are useful in the construction, in ordering models, and in the relation between cuttings of different formulas.

The next example, illustrating such a construction, shows how our general definition via cuttings can capture some explanatory relations defined in the literature.

Example 3. Abduction via semantic tableau [14] and resolution [40] generates a cutting, and an explanatory relation. We illustrate this fact for abduction via semantic tableau in the framework of propositional logic.[5]

Semantic tableaux are used as refutation systems. Let S be a set of propositional formulas. The tableau expansion rules are as follows:

$$\neg - rules: \quad \frac{S \cup \{\neg\neg\varphi\}}{S \cup \{\varphi\}} \qquad \frac{S \cup \{\neg\bot\}}{S \cup \{\top\}}$$

$$\alpha - rules: \quad \frac{S \cup \{\varphi_1 \wedge \varphi_2\}}{S \cup \{\varphi_1, \varphi_2\}} \qquad \frac{S \cup \{\neg(\varphi_1 \Rightarrow \varphi_2)\}}{S \cup \{\varphi_1, \neg\varphi_2\}} \qquad \frac{S \cup \{\neg(\varphi_1 \vee \varphi_2)\}}{S \cup \{\neg\varphi_1, \neg\varphi_2\}}$$

$$\beta - rules: \quad \frac{S \cup \{\varphi_1 \vee \varphi_2\}}{\{S \cup \{\varphi_1\}, S \cup \{\varphi_2\}\}} \qquad \frac{S \cup \{\varphi_1 \Rightarrow \varphi_2\}}{\{S \cup \{\neg\varphi_1\}, S \cup \{\varphi_2\}\}} \qquad \frac{S \cup \{\neg(\varphi_1 \wedge \varphi_2)\}}{\{S \cup \{\neg\varphi_1\}, S \cup \{\neg\varphi_2\}\}}$$

A tableau $\mathcal{T}$ is then a sequence of sets of sets of formulas $(\Gamma_1, \ldots, \Gamma_n, \ldots)$ such that, for every i, Γ_{i+1} is obtained from Γ_i by the application of a tableau expansion rule on a formula of a set S in Γ_i. At each step i, every set S in Γ_i which contains both p and $\neg p$ for some propositional variable p is removed from Γ_i.

A formula φ is a theorem of a KB T if there exists a finite sequence $(\Gamma_1, \ldots, \Gamma_n)$ such that $\Gamma_1 = \{T \cup \{\neg\varphi\}\}$ and $\Gamma_n = \emptyset$. As an example, let us show that a is a theorem of $\{a \wedge c, a \Rightarrow b\}$. The tableau method provides the finite sequence $\Gamma_1 = \{\{a \wedge c, a \Rightarrow b, \neg a\}\}, \Gamma_2 = \{\{a, c, a \Rightarrow b, \neg a\}\}$, using α-rules. The set Γ_2 contains a unique set, with both a and $\neg a$, which is then removed, and Γ_2 becomes empty.

Let us observe that the tableau expansion rules break propositional formulas on their main Boolean connectives. Hence, tableaux are necessarily finite, and two cases can occur:

(1) the last set Γ_n of the sequence is empty, and we have that $T \models \varphi$; or
(2) every S in Γ_n only contains literals but no literal has its negation in S.

Following [14], if g is any consistent choice function for the elements of Γ_n, i.e. for $\Gamma_n = \{S_{n1}, \ldots, S_{nm_n}\}$, $g(S_{ni}) \in S_{ni}$, then if $\psi = \neg g(S_{n1}) \wedge \ldots \wedge \neg g(S_{nm_n})$ is consistent with T, then ψ is an explanation of φ for T (ψ is even the minimal one according to the definition of minimality given in [14]).

We now show that the way the tableau $\mathcal{T}$ is generated in [14] defines a cutting $\mathcal{C}_\mathcal{T}$. Before defining the cutting $\mathcal{C}_\mathcal{T}$, let us introduce some useful notions. Let $\mathcal{T} = (\Gamma_1, \ldots, \Gamma_n)$ be a tableau for $T \cup \{\neg\varphi\}$ such that $\Gamma_i = \{S_{i1}, \ldots, S_{im_i}\}$. For every j,

[5] Note that semantic tableau methods have been extended to modal logic [6,15], first-order logic [33,39], DL [28], etc., and in the same way we would be able to generate a cutting from them.

$1 \leq j \leq m_i$, let us denote ψ_{ij} the disjunction of the negation of all the literals $l \in S_{ij}$, i.e. $\psi_{ij} = \bigvee\{\neg l \mid l : literal \text{ and } l \in S_{ij}\}$. Then, let us set $\psi_i = \bigwedge_{1 \leq j \leq m_i} \psi_{ij}$. We can define the cutting $\mathcal{C}_\mathcal{T}$ as follows:

$$\mathcal{C}_\mathcal{T} = \{Mod(T \cup \{\varphi\})\} \cup (\cup_{1 \leq i \leq n}\{Mod(\psi_i)\})$$

Obviously we have $Mod(T \cup \{\varphi\}) \in \mathcal{C}_\mathcal{T}$ and $Triv \notin \mathcal{C}_\mathcal{T}$. Moreover, for any i, $Mod(\psi_i) \subseteq Mod(T \cup \{\varphi\})$, hence $\mathcal{C}_\mathcal{T} \subseteq \mathcal{P}(Mod(T \cup \{\varphi\}))$. It is not difficult to show that for every i, $1 \leq i \leq n$, $Mod(\psi_{i+1}) \subseteq Mod(\psi_i) \subseteq Mod(T \cup \{\varphi\})$. Moreover, the tableau $\mathcal{T}$ is finite, which completes the proof that $\mathcal{C}_\mathcal{T}$ is a cutting.

Let us illustrate this construction on an example. Let $T = \{f \Rightarrow m, t \vee s, r \Rightarrow m\}$ be the KB and let $\varphi = m$ be the observation. The tableau method applied to $T \cup \{\neg\varphi\}$ generates four sets $\Gamma_1, \ldots, \Gamma_4$ where:

- $\Gamma_1 = \{\{f \Rightarrow m, t \vee s, r \Rightarrow m, \neg m\}\}$;
- $\Gamma_2 = \{\{\neg f, t \vee s, r \Rightarrow m, \neg m\}\}$;
- $\Gamma_3 = \{\{\neg f, t, r \Rightarrow m, \neg m\}, \{\neg f, s, r \Rightarrow m, \neg m\}\}$;
- $\Gamma_4 = \{\{\neg f, t, \neg r, \neg m\}, \{\neg f, s, \neg r, \neg m\}\}$.

This leads to the following formulas $\psi_1, \ldots, \psi_4$:

- $\psi_1 = m$;
- $\psi_2 = f \vee m$;
- $\psi_3 = (f \vee m \vee \neg t) \wedge (f \vee m \vee \neg s)$;
- $\psi_4 = (f \vee m \vee \neg t \vee r) \wedge (f \vee m \vee \neg s \vee r)$.

A consistent choice satisfying minimality is for instance $f \vee r$.

It is interesting to note that there is an alternative way of looking at $\rhd_\mathcal{C}$. The descending chains to obtain the minimal element $\mathbb{M}$ provide a method to order the models of $Mod(T \cup \{\varphi\})$. This ordering, expressing when a model is preferred to another one, corresponds to the idea of "preferred" explanations according to a minimality criterion, the minimality being represented by the most central models as in Fig. 1.

Definition 8 *(Relation on models).* Let T be a KB and let φ be a formula such that $Mod(T \cup \{\varphi\}) \neq Triv$. Let $\mathcal{C}_\varphi$ be a cutting for T and φ. Let us define $\preceq_{\mathcal{C}_\varphi} \subseteq Mod \times Mod$ as follows:

$$\mathcal{M} \preceq_{\mathcal{C}_\varphi} \mathcal{M}' \Longleftrightarrow \begin{cases} \exists C \subseteq \mathcal{C}_\varphi, s.t. C \text{ is a maximal chain} \\ \forall \mathbb{M} \in C, \mathcal{M}' \in \mathbb{M} \Rightarrow \mathcal{M} \in \mathbb{M} \end{cases} \quad (1)$$

Let $\mathbb{M} \subseteq Mod$ and $\preceq$ be a binary relation over $\mathbb{M}$. We define $\prec$ as $\mathcal{M} \prec \mathcal{M}'$ if and only if $\mathcal{M} \preceq \mathcal{M}'$ and $\mathcal{M}' \not\preceq \mathcal{M}$. We also define $Min(\mathbb{M}, \preceq) = \{\mathcal{M} \in \mathbb{M} \mid \forall \mathcal{M}' \in \mathbb{M}, \mathcal{M}' \not\prec \mathcal{M}\}$. Note that the relation $\preceq_{\mathcal{C}_\varphi}$ is reflexive, but not necessarily transitive (hence it is not a pre-order).

Theorem 1. *Let $\mathcal{C}_\varphi$ be the cutting for a KB T and a formula φ used in the definition of $\rhd_\mathcal{C}$. For any $\psi \in Sen$, the following equivalence holds:*

$$\varphi \rhd_\mathcal{C} \psi \Longleftrightarrow \begin{cases} Mod(T \cup \{\psi\}) \setminus Triv \neq \emptyset, \text{ and} \\ Mod(T \cup \{\psi\}) \setminus Triv \subseteq Min(Mod(T \cup \{\varphi\}) \setminus Triv, \preceq_{\mathcal{C}_\varphi}) \end{cases}$$

Proof. ($\Rightarrow$) By definition of $\rhd_\mathcal{C}$, we have $Mod(T \cup \{\psi\}) \setminus Triv \neq \emptyset$. Let us suppose $\mathcal{M} \in Mod(T \cup \{\psi\}) \setminus Triv$. By the definition of $\rhd_\mathcal{C}$, the statement $\varphi \rhd_\mathcal{C} \psi$ means that there exists $\mathbb{M} \in Min(\mathcal{C}_\varphi)$ such that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}$. As $(\mathcal{C}_\varphi, \subseteq)$ satisfies the Hausdorff maximal principle, there exists a maximal chain C, the least element of which is $\mathbb{M}$. Hence, by the definition of $\preceq_{\mathcal{C}_\varphi}$, for every $\mathbb{M}' \in C$, we have that:

(1) for every $\mathcal{M}' \in \mathbb{M}$, $\mathcal{M} \preceq_{\mathcal{C}_\varphi} \mathcal{M}'$ and $\mathcal{M}' \preceq_{\mathcal{C}_\varphi} \mathcal{M}$, and
(2) for every $\mathcal{M}' \in \mathbb{M}' \setminus \mathbb{M}$, $\mathcal{M} \prec_{\mathcal{C}_\varphi} \mathcal{M}'$.

This proves that $\mathcal{M} \in Min(Mod(T \cup \{\varphi\}) \setminus Triv, \preceq_{\mathcal{C}_\varphi})$.

($\Leftarrow$) Let us suppose that $\varphi \not\rhd_\mathcal{C} \psi$. This means that either $Mod(T \cup \{\psi\}) = Triv$ and in this case the conclusion is obvious, or there does not exist a minimal element $\mathbb{M} \in Min(\mathcal{C}_\varphi)$ such that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}$. Let $\mathbb{M}$ be the least element (for inclusion) of $\mathcal{C}_\varphi$ such that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}$. This least element $\mathbb{M}$ exists because $\mathcal{C}_\varphi$ contains $Mod(T \cup \{\varphi\})$ and $(\mathcal{C}_\varphi, \subseteq)$ is well-founded. As $(\mathcal{C}_\varphi, \subseteq)$ satisfies the Hausdorff maximal principle, there exists a maximal chain C which contains $\mathbb{M}$, and therefore $\mathbb{M}$ cannot be the least element of C. Therefore, there exist some models $\mathcal{M}'$ which belong to some elements $\mathbb{M}'$ in C such that $Mod(T \cup \{\psi\}) \not\subseteq \mathbb{M}'$, and therefore, for some models $\mathcal{M} \in Mod(T \cup \{\psi\}) \setminus Triv$, we have $\mathcal{M}' \prec_{\mathcal{C}_\varphi} \mathcal{M}$. □

The explanatory relation $\rhd_{\mathcal{C}}$ satisfies a number of logical properties. Most of these properties are (rationality) postulates defined in [35] up to some adaptations. Let us recall them, adapted to the satisfaction system context, for any KB T, explanatory relation $\rhd$ for T and formulas $\varphi, \varphi', \psi \in Sen$:

LLE	$\dfrac{\varphi \equiv_T \varphi' \quad \varphi \rhd \psi}{\varphi' \rhd \psi}$
RLE	$\dfrac{\psi \equiv_T \psi' \quad \varphi \rhd \psi}{\varphi \rhd \psi'}$
E-CM	$\dfrac{\varphi \rhd \psi \quad T \cup \{\psi\} \models \varphi'}{\varphi \wedge \varphi' \rhd \psi}$
E-C-Cut	$\dfrac{\varphi \wedge \varphi' \rhd \psi \quad \forall \psi' (\varphi \rhd \psi' \Rightarrow T \cup \{\psi'\} \models \varphi')}{\varphi \rhd \psi}$
E-R-Cut	$\dfrac{\varphi \wedge \varphi' \rhd \psi \quad \exists \psi' (\varphi \rhd \psi' \text{ and } T \cup \{\psi'\} \models \varphi')}{\varphi \rhd \psi}$
LOR	$\dfrac{\varphi \rhd \psi \quad \varphi' \rhd \psi}{\varphi \vee \varphi' \rhd \psi}$
E-DR	$\dfrac{\varphi \rhd \psi \quad \varphi' \rhd \psi'}{\varphi \vee \varphi' \rhd \psi \text{ or } \varphi \vee \varphi' \rhd \psi'}$
ROR	$\dfrac{\varphi \rhd \psi \quad \varphi \rhd \psi'}{\varphi \rhd \psi \vee \psi'}$
RS	$\dfrac{\varphi \rhd \psi \quad K \cup \{\psi'\} \models \psi \quad Mod(T \cup \{\psi'\}) \neq Triv}{\varphi \rhd \psi'}$
E-Reflexivity	$\dfrac{\varphi \rhd \psi}{\psi \rhd \psi}$
E-Con	$Mod(T \cup \{\varphi\}) \neq Triv \iff \exists \psi, \varphi \rhd \psi$

Now, we will show that, with an appropriate structure on the set of cuttings $\mathcal{C}$, adding a limited set of rather intuitive stability and monotony requirements, we can get strong results on the explanatory relation $\rhd_{\mathcal{C}}$, according to the above postulates. Recall that $\mathcal{C}$ is defined by choosing a cutting $\mathcal{C}_\varphi$ for each φ in Sen. A first requirement is that for every φ, φ' we have:

$$\text{If } \varphi \equiv_T \varphi', \text{ then } \mathcal{C}_\varphi = \mathcal{C}_{\varphi'} \tag{2}$$

This will be directly used in Property (1) of the following Theorem.

Theorem 2. *Let $\mathcal{R}$ be a satisfaction system, T a KB, $\mathcal{C}$ a set of cuttings and $\rhd_{\mathcal{C}}$ the explanatory relation based on cuttings of Definition 7. The following properties are satisfied, for every $\varphi, \varphi', \psi, \psi'$:*

(1) *Assume that $\mathcal{C}$ satisfies Equation (2). If $\varphi \equiv_T \varphi'$, $\psi \equiv_T \psi'$ and $\varphi \rhd_{\mathcal{C}} \psi$, then $\varphi' \rhd_{\mathcal{C}} \psi'$.*
(2) *If $\varphi \rhd_{\mathcal{C}} \psi$ and $T \cup \{\psi'\} \models \psi$ with $Mod(T \cup \{\psi'\}) \neq Triv$, then $\varphi \rhd_{\mathcal{C}} \psi'$.*
(3) *$\psi \in Expla_T(\varphi)$ iff there exists a relation $\rhd_{\mathcal{C}}$ based on cuttings such that $\varphi \rhd_{\mathcal{C}} \psi$.*
(4) *If $\mathcal{R}$ is finitely axiomatizable for every $\mathbb{M} \subseteq Mod$ and has conjunction, then for every cutting $\mathcal{C}_\varphi$, we have that $Expla_T(\varphi) \neq \emptyset$ and $\exists \psi \in Sen, \varphi \rhd_{\mathcal{C}} \psi$, where $\rhd_{\mathcal{C}}$ is a relation based on cuttings such that the cutting associated with φ is $\mathcal{C}_\varphi$.*

Proof. (1) The first property is obviously satisfied because $\varphi \equiv_T \varphi'$ and $\psi \equiv_T \psi'$ mean that $Mod(T \cup \{\varphi\}) = Mod(T \cup \{\varphi'\})$ and $Mod(T \cup \{\psi\}) = Mod(T \cup \{\psi'\})$ and, by assumption, $\mathcal{C}_\varphi = \mathcal{C}_{\varphi'}$.
(2) $\varphi \rhd_{\mathcal{C}} \psi$ means that there exists $\mathbb{M} \in Min(\mathcal{C}_\varphi)$ such that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}$, and $Mod(T \cup \{\varphi\}) \neq Triv$. As $Mod(T \cup \{\psi'\}) \subseteq Mod(T \cup \{\psi\})$ (hence $Mod(T \cup \{\psi'\}) \subseteq \mathbb{M}$), and $Mod(T \cup \{\psi'\}) \neq Triv$, we can deduce that $\varphi \rhd_{\mathcal{C}} \psi'$.
(3) The "if part" is obvious. To prove the "only if part", let us notice that for every $\psi \in Expla_T(\varphi)$, we can build in $\mathcal{P}(Mod(T \cup \{\varphi\})$ a saturated chain $\mathcal{C}_\varphi$ starting at $Mod(T \cup \{\psi\})$. As $\psi \in Expla_T(\varphi)$, this saturated chain satisfies all the conditions of Definition 6, and therefore it is a cutting for T and φ. By Remark 2, we can define the relation $\rhd_{\mathcal{C}}$ based on cuttings such that the cutting corresponding to φ is precisely $\mathcal{C}_\varphi$. By construction of $\mathcal{C}_\varphi$, it is clear that $\varphi \rhd_{\mathcal{C}} \psi$.
(4) Again by Remark 2, it makes sense to consider $\rhd_{\mathcal{C}}$ as the explanatory relation defined by a family of cuttings in which the one associated with φ is $\mathcal{C}_\varphi$. Let $\mathbb{M} \in Min(\mathcal{C}_\varphi)$. As $\mathcal{R}$ is finitely axiomatizable, there exists a finite KB T' such that $Mod(T') = \mathbb{M}$. Let us set $\psi = \bigwedge_{\varphi' \in T'} \varphi'$. We obviously have that $T \cup T'$ is consistent, hence $Mod(T \cup \{\psi\}) \neq Triv$, and we can conclude that $\varphi \rhd_{\mathcal{C}} \psi$. □

It is interesting to note that Property (1) generalizes to satisfaction systems the properties LLE and RLE of [35]. Similarly, Property (2) corresponds to RS, and Properties (3) and (4) to E-Con.

If $\mathcal{R}$ also has Boolean connectives in $\{\wedge, \vee, \Rightarrow\}$, the explanatory relation $\rhd_{\mathcal{C}}$ satisfies additional logical properties.

Theorem 3. *Let $\mathcal{R}$ be a satisfaction system with conjunction, disjunction and implication. Let T be a KB, and $\mathcal{C} = \{\mathcal{C}_\varphi \mid \varphi \in Sen, Mod(T \cup \{\varphi\}) \setminus Triv \neq \emptyset\}$ a set of cuttings. Then, the following properties are satisfied, for every $\varphi, \varphi', \psi, \psi'$:*

(5) *If* $\varphi \rhd_{\mathcal{C}} \psi$ *and* $Mod(T \cup \{\psi \wedge \psi'\}) \neq Triv$, *then* $\varphi \rhd_{\mathcal{C}} \psi \wedge \psi'$.
(6) *If* $\varphi \rhd_{\mathcal{C}} \psi \wedge \psi'$ *and* $T \cup \{\psi\} \models \psi'$, *then* $\varphi \rhd_{\mathcal{C}} \psi$.
(7) *If* $\varphi \rhd_{\mathcal{C}} \psi \vee \psi'$ *and* $Mod(T \cup \{\psi\}) \neq Triv$, *then* $\varphi \rhd_{\mathcal{C}} \psi$.
(8) *For every cutting* $\mathcal{C}_\varphi$ *such that the ordering relation* $\preceq_{\mathcal{C}_\varphi}$ *is total, if* $\varphi \rhd_{\mathcal{C}} \psi$ *and* $\varphi \rhd_{\mathcal{C}} \psi'$, *then* $\varphi \rhd_{\mathcal{C}} \psi \vee \psi'$.
(9) *If* $\varphi \rhd_{\mathcal{C}} \psi$, *then* $\psi \rhd_{\mathcal{C}} \psi$, *for* $\mathcal{C}_\psi = \{\mathbb{M} \cap Mod(\psi) \mid \mathbb{M} \in \mathcal{C}_\varphi\}$.

Proof. (5) By hypothesis, there exists $\mathbb{M} \in Min(\mathcal{C}_\varphi)$ such that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}$. Obviously, we have that $Mod(T \cup \{\psi \wedge \psi'\}) \subseteq Mod(T \cup \{\psi\})$, and as $T \cup \{\psi \wedge \psi'\})$ is consistent, we can deduce that $\varphi \rhd_{\mathcal{C}} \psi \wedge \psi'$.
(6) By hypothesis, there exists $\mathbb{M} \in Min(\mathcal{C}_\varphi)$ such that $Mod(T \cup \{\psi \wedge \psi'\}) \subseteq \mathbb{M}$. Since $T \cup \{\psi\} \models \psi'$, we have that $Mod(T \cup \{\psi \wedge \psi'\}) = Mod(T \cup \{\psi\})$, and we can conclude that $\varphi \rhd_{\mathcal{C}} \psi$.
(7) This property is a direct consequence of the fact that $Mod(\psi) \subseteq Mod(\psi \vee \psi')$.
(8) By the hypothesis that $\preceq_{\mathcal{C}_\varphi}$ is total, the poset $(\mathcal{C}_\varphi, \subseteq)$ contains a unique maximal chain C. Hence, both $Mod(T \cup \{\psi\})$ and $Mod(T \cup \{\psi'\})$ are included in the unique minimal element $\mathbb{M}$ of C. Obviously, we have that $Mod(T \cup \{\psi \vee \psi'\}) \subseteq \mathbb{M}$, and we can conclude that $\varphi \rhd_{\mathcal{C}} \psi \vee \psi'$.
(9) This property is obviously satisfied. □

Properties (8) and (9) are extensions of the postulates ROR and E-Reflexivity defined in [35]. Properties (5) and (7) are revisited forms of the postulate RS, adapted to satisfaction systems and explanations based on cuttings.

Lemma 1. *If* $\mathcal{C}_\varphi$ *is a cutting for* T *and* φ, $Mod(\varphi') \subseteq Mod(\varphi)$ *and* $Mod(T \cup \{\varphi'\}) \neq Triv$ *then* $\mathcal{C}_{\varphi'} = \{\mathbb{M} \cap Mod(\varphi') \mid \mathbb{M} \in \mathcal{C}_\varphi, \mathbb{M} \cap Mod(\varphi') \neq Triv\}$ *is a cutting for* T *and* φ'.

Proof. Let $\mathbb{M} \in \mathcal{C}_\varphi$ such that $\mathbb{M} \cap Mod(\varphi') \neq Triv$. By hypothesis, we have that $Triv \subset \mathbb{M} \cap Mod(\varphi')$. Note that $Mod(T \cup \{\varphi\}) \in \mathcal{C}_\varphi$, $Mod(T \cup \{\varphi\}) \cap Mod(\varphi') = Mod(T \cup \{\varphi'\})$, and $Mod(T \cup \{\varphi'\}) \neq Triv$. Therefore $Mod(T \cup \{\varphi'\}) \in \mathcal{C}_{\varphi'}$. Moreover, $\mathcal{C}_{\varphi'}$ is obviously closed under set-theoretical union. Finally, it is clear that if $\mathcal{C}_{\varphi'}$ is not well-founded, then neither is $\mathcal{C}_\varphi$. □

Theorem 4. *Let* $\mathcal{R}$ *be a satisfaction system with conjunction, disjunction and implication. Let* T *be a KB, and* $\mathcal{C} = \{\mathcal{C}_\varphi \mid \varphi \in Sen, Mod(T \cup \{\varphi\}) \setminus Triv \neq \emptyset\}$ *a set of cuttings satisfying the following condition, called Condition SUB-CUT:*

$$\forall \varphi, \varphi' \in Sen, Mod(\varphi') \subseteq Mod(\varphi),\ \mathcal{C}_{\varphi'}, \mathcal{C}_\varphi \in \mathcal{C} \Longrightarrow \mathcal{C}_{\varphi'} = \{\mathbb{M} \cap Mod(\varphi') \mid \mathbb{M} \in \mathcal{C}_\varphi, \mathbb{M} \cap Mod(\varphi') \neq Triv\}$$

The following properties are satisfied, for every $\varphi, \varphi', \psi, \psi'$:

(10) *If* $\varphi \rhd_{\mathcal{C}} \psi$ *and* $T \cup \{\psi\} \models \varphi'$, *then* $\varphi \wedge \varphi' \rhd_{\mathcal{C}} \psi$.
(11) *Assume that for every* $\varphi \in Sen$, *all* $\mathbb{M} \in \mathcal{C}_\varphi$ *are finitely axiomatizable. If* $\varphi \wedge \varphi' \rhd_{\mathcal{C}} \psi$ *and for every* ψ' *such that* $\varphi \rhd_{\mathcal{C}} \psi'$, *we have* $T \cup \{\psi'\} \models \varphi'$, *then* $\varphi \rhd_{\mathcal{C}} \psi$.
(12) *If* $\varphi \vee \varphi' \rhd_{\mathcal{C}} \psi$ *and* $T \cup \{\psi\} \models \varphi$, *then* $\varphi \rhd_{\mathcal{C}} \psi$.
(13) *If* $(\varphi \Rightarrow \varphi') \rhd_{\mathcal{C}} \psi$ *and* $T \cup \{\psi\} \models \varphi$, *then* $\varphi' \rhd_{\mathcal{C}} \psi$.

Proof. (10) By hypothesis, there exists $\mathbb{M} \in Min(\mathcal{C}_\varphi)$ such that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}$ and $Mod(T \cup \{\psi\}) \neq Triv$. As we have further $Mod(T \cup \{\psi\}) \subseteq Mod(\varphi')$, we have that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M} \cap Mod(\varphi')$. Note also that $\mathbb{M} \cap Mod(\varphi') = \mathbb{M} \cap Mod(\varphi' \wedge \varphi)$. Then, by Condition SUB-CUT, $\mathbb{M} \cap Mod(\varphi' \wedge \varphi) \in \mathcal{C}_{\varphi \wedge \varphi'}$, and due to the fact that $\mathbb{M} \in Min(\mathcal{C}_\varphi)$, necessarily $\mathbb{M} \cap Mod(\varphi' \wedge \varphi)$ is a minimal element of $\mathcal{C}_{\varphi \wedge \varphi'}$. Therefore, we can conclude that $\varphi \wedge \varphi' \rhd_{\mathcal{C}} \psi$.
(11) Take $\mathbb{M}' \in Min(\mathcal{C}_{\varphi \wedge \varphi'})$ such that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}'$. By Condition SUB-CUT, $\mathbb{M}' = \mathbb{M} \cap Mod(\varphi \wedge \varphi')$ with $\mathbb{M} \in \mathcal{C}_\varphi$. Let $\mathbb{M}''$ be in $Min(\mathcal{C}_\varphi)$ such that $\mathbb{M}'' \subseteq \mathbb{M}$. Since every $\mathbb{M}$ is finitely aziomatizable, there exists ψ' such that $Mod(\psi') = \mathbb{M}''$. Note that $\mathbb{M}'' \subseteq Mod(T)$, thus $Mod(T \cup \{\varphi'\}) = \mathbb{M}''$ and therefore, $\varphi \rhd_{\mathcal{C}} \psi'$. Then, by hypothesis, $T \cup \{\psi'\} \models \varphi'$, i.e., $\mathbb{M}'' \subseteq Mod(\varphi')$. Note that $\mathbb{M}'' \cap Mod(\varphi \wedge \varphi') \subseteq \mathbb{M} \cap Mod(\varphi \wedge \varphi')$. But the right member of the inclusion is minimal in $\mathcal{C}_{\varphi \wedge \varphi'}$, then we have the equality. Note also that $\mathbb{M}'' = \mathbb{M}'' \cap Mod(\varphi \wedge \varphi')$. Thus, $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}''$, i.e., $\varphi \rhd_{\mathcal{C}} \psi$.
(12) Let $\mathbb{M} \in Min(\mathcal{C}_{\varphi \vee \varphi'})$ such that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}$. Since $Mod(T \cup \{\varphi\}) \neq Triv$ and $Mod(T \cup \{\varphi\}) \subseteq Mod(\varphi)$, necessarily $\mathbb{M} \cap Mod(\varphi) \neq Triv$. Thus, by Condition SUB-CUT, $\mathbb{M} \cap Mod(\varphi) \in \mathcal{C}_\varphi$. Moreover, due to the fact that $\mathbb{M} \in Min(\mathcal{C}_{\varphi \vee \varphi'})$, we have that $\mathbb{M} \cap Mod(\varphi) \in Min(\mathcal{C}_\varphi)$. Finally, since $Mod(T \cup \{\psi\}) \subseteq \mathbb{M} \cap Mod(\varphi)$, we can conclude that $\varphi \rhd_{\mathcal{C}} \psi$.
(13) Let $\mathbb{M} \in Min(\mathcal{C}_{\varphi \Rightarrow \varphi'})$ such that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M}$. Since $Mod(T \cup \{\psi\}) \subseteq Mod(\varphi)$, we have that $Mod(T \cup \{\psi\}) \subseteq \mathbb{M} \cap Mod(\varphi')$. Note that $Mod(T \cup \{\psi\}) \neq Triv$, thus $\mathbb{M} \cap Mod(\varphi') \neq Triv$. Therefore, by Condition SUB-CUT, $\mathbb{M} \cap Mod(\varphi') \in \mathcal{C}_{\varphi'}$ and, due to the fact that $\mathbb{M} \in Min(\mathcal{C}_{\varphi \Rightarrow \varphi'})$, we have that $\mathbb{M} \cap Mod(\varphi') \in Min(\mathcal{C}_{\varphi'})$, and we can conclude that $\varphi' \rhd_{\mathcal{C}} \psi$. □

Properties (10) is an extension of the postulate E-CM defined in [35]. Property (11) is a revisited form of the postulate E-C-Cut, adapted to satisfaction systems and explanations based on cuttings.

The following two results are easy to prove, and therefore we omit the proof.

Lemma 2. *Let* $\mathcal{C}_\varphi$ *and* $\mathcal{C}_{\varphi'}$ *be cuttings for* T *and* φ *(respectively* φ'*). Then the set* $\mathcal{C}_\varphi \bigoplus \mathcal{C}_{\varphi'} = \{A \cup B \mid A \in \mathcal{C}_\varphi \text{ and } B \in \mathcal{C}_{\varphi'}\}$ *is a cutting for* T *and* $\varphi \vee \varphi'$.

Corollary 1. *Suppose that $\rhd_{\mathcal{C}}$ is an explanatory relation defined on cuttings such that $\mathcal{C}_\varphi$ and $\mathcal{C}_{\varphi'}$ are the cuttings for T and φ and for T and φ' respectively. Furthermore, suppose that the cutting for $\varphi \vee \varphi'$ is precisely $\mathcal{C}_\varphi \bigoplus \mathcal{C}_{\varphi'}$. If $\varphi \rhd_{\mathcal{C}} \psi$ or $\varphi' \rhd_{\mathcal{C}} \psi$, then $\varphi \vee \varphi' \rhd_{\mathcal{C}} \psi$.*

This result is a stronger version of LOR.

4. Cutting based on retraction

In this section, we introduce some more constraints on cuttings, to be able to propose concrete examples in various logics in Section 5. The idea is to define particular cuttings from "retractions", that consist in transforming any KB T into a new consistent one T' such that $Mod(T') \subseteq Mod(T)$. This notion of retraction draws inspiration from Bloch & al.'s work in [8–10] on Morpho-Logics where some retractions have been defined based on erosions from mathematical morphology [7] (see Section 5). Here, we propose to generalize this notion in the framework of satisfaction systems, and a retraction is defined on Sen as follows.

Definition 9 *(Retraction).* A **retraction** is a mapping $\kappa : Sen \to Sen$ satisfying, for every $\varphi \in Sen$ such that $Mod(\varphi) \neq Triv$, the two following properties:

- **Anti-extensivity:** $Mod(\kappa(\varphi)) \subseteq Mod(\varphi)$.
- **Vacuum:** $\exists k \in \mathbb{N}, Mod(\kappa^k(\varphi)) = Triv$ where κ^0 is the identity mapping, and for all $k > 0$, $\kappa^k(\varphi) = \kappa(\kappa^{k-1}(\varphi))$.

The condition for φ not to be a tautology in Definition 9 allows us to eliminate the trivial case where $(\bigwedge T) \wedge \varphi$ is a tautology for a KB T and an observation φ, and in this case φ would not deserve any explanation from T.

Example 4. Many examples of retractions can be defined in **PL**. Here, we propose to define retractions from the tableau expansion rules given in Example 3. In Section 5.1, we will study another retraction for **PL** but based on erosions from mathematical morphology.

Let us recall that the tableau expansion rules break propositional formulas on their main Boolean connectives, and only the β-rules require a choice. Hence, given a choice such as for instance choosing at each time the left element of the formula (i.e. the formula φ_1 in β-rules), we start by inductively defining a mapping $h : Sen \to Sen$ as follows:

(1) $h(p) = p$ for every $p \in \Sigma$;
(2) $h(\neg\neg\varphi) = h(\varphi)$;
(3) $h(\neg\neg\bot) = \top$;
(4) $h(\varphi_1 \wedge \varphi_2) = h(\varphi_1) \wedge h(\varphi_2)$;
(5) $h(\neg(\varphi_1 \Rightarrow \varphi_2)) = h(\varphi_1) \wedge h(\neg\varphi_2)$;
(6) $h(\neg(\varphi_1 \vee \varphi_2)) = h(\neg\varphi_1) \wedge h(\neg\varphi_2)$;
(7) $h(\varphi_1 \vee \varphi_2) = h(\varphi_1)$;
(8) $h(\varphi_1 \Rightarrow \varphi_2) = h(\neg\varphi_1)$;
(9) $h(\neg(\varphi_1 \wedge \varphi_2)) = h(\neg\varphi_1)$.

We could have just as easily defined the mapping h as follows: the first six cases are identical, and

- $h(\varphi_1 \vee \varphi_2) = h(\varphi_2)$ or $h(\varphi_1 \vee \varphi_2) = h(\varphi_1) \wedge h(\varphi_2)$;
- $h(\varphi_1 \Rightarrow \varphi_2) = h(\varphi_2)$ or $h(\varphi_1 \Rightarrow \varphi_2) = h(\varphi_1) \wedge h(\varphi_2)$;
- $h(\neg(\varphi_1 \wedge \varphi_2)) = h(\neg\varphi_2)$ or $h(\neg(\varphi_1 \wedge \varphi_2)) = h(\neg\varphi_1) \wedge h(\neg\varphi_2)$.

Hence, each application of the mapping h corresponds to a step of a path in the tableau. By structural induction on φ, it is easy to show that $Mod(h(\varphi)) \subseteq Mod(\varphi)$, hence h verifies the anti-extensivity property. Moreover, it is also obvious to show that $h(h(\varphi)) = h(\varphi)$, and, except if φ is an antilogy (i.e. a contradiction), we cannot have $Mod(h^k(\varphi)) = Triv$ for some k. This means that h does not satisfy the vacuum property. Hence, h is not a retraction. Now it is quite obvious to define a retraction from h. Indeed, given a mapping h defined as previously, let us define the retraction $\kappa_h : Sen \to Sen$ as follows:

$$\kappa_h(\varphi) = \begin{cases} \bot & \text{if } h(\varphi) = \varphi \\ h(\varphi) & \text{otherwise} \end{cases}$$

It is not difficult to show that κ_h is a retraction.

Here, we introduce two cuttings based on retraction: $\mathcal{C}_{lcr}$ (last consistent retraction) and $\mathcal{C}_{lnr}$ (last non-trivial retraction). In $\mathcal{C}_{lcr}$ (respectively in $\mathcal{C}_{lnr}$), we define a unique sequence of ordered models (cf. Proposition 3) which approximates the *most central part* of T (respectively of $T \cup \{\varphi\}$).

Definition 10 *(Cuttings based on retraction).* Let κ be a retraction. Let $T \subseteq Sen$ be a KB and φ be a sentence such that $T \cup \{\varphi\}$ is consistent. Let us define the two subsets of $\mathcal{P}(Mod(T \cup \{\varphi\})$ as follows:

$$\mathcal{C}^{\varphi}_{lcr} = \{Mod(\kappa^k(\bigwedge T) \wedge \varphi) \mid k \in \mathbb{N}, Mod(\kappa^k(\bigwedge T) \wedge \varphi) \neq Triv\} \tag{3}$$

$$\mathcal{C}^{\varphi}_{lnr} = \{Mod(\kappa^k(\bigwedge T \wedge \varphi)) \mid k \in \mathbb{N}, Mod(\kappa^k(\bigwedge T \wedge \varphi)) \neq Triv\} \tag{4}$$

where $\bigwedge T = \varphi_1 \wedge \ldots \wedge \varphi_n$ if $T = \{\varphi_1, \ldots, \varphi_n\}$.

Proposition 2. *$\mathcal{C}^{\varphi}_{lcr}$ and $\mathcal{C}^{\varphi}_{lnr}$ as defined in Equations (3) and (4) are cuttings for T and φ.*

Proof. Let us observe that in $\mathcal{C}^{\varphi}_{lcr}$ (respectively in $\mathcal{C}^{\varphi}_{lnr}$) we have a unique maximal chain of finite size the least element of which is $Mod(\kappa^n(\bigwedge T) \wedge \varphi)$ (respectively $Mod(\kappa^n(\bigwedge T \wedge \varphi))$ where $n = \sup\{k \in \mathbb{N} \mid Mod(\kappa^n(\bigwedge T) \wedge \varphi) \neq Triv\}$ (respectively $Mod(\kappa^n(\bigwedge T \wedge \varphi)) \neq Triv$) (by the vacuum property such a n exists). Obviously, both sets contain $Mod(T \cup \{\varphi\})$ (obtained for $k = 0$), are closed under set-theoretical inclusion and are well-founded. □

Proposition 3. *Both $\preceq_{\mathcal{C}^{\varphi}_{lcr}}$ and $\preceq_{\mathcal{C}^{\varphi}_{lnr}}$ derived from the cuttings defined in Equations (3) and (4) as in Equation (1) are total pre-orders.*

Proof. This is a direct consequence of the fact that both $\mathcal{C}^{\varphi}_{lcr}$ and $\mathcal{C}^{\varphi}_{lnr}$ have a unique maximal chain. □

Let us denote $\mathcal{C}_{lcr} = \{\mathcal{C}^{\varphi}_{lcr} \mid \varphi \in Sen\}$ and $\mathcal{C}_{lnr} = \{\mathcal{C}^{\varphi}_{lnr} \mid \varphi \in Sen\}$. Following Definition 7, these two sets of cuttings give rise to two explanatory relations defined as follows:

$$\varphi \rhd_{\mathcal{C}_{lcr}} \psi \iff \begin{cases} Mod(T \cup \{\psi\}) \neq Triv, \text{ and} \\ Mod(T \cup \{\psi\}) \subseteq Mod(\kappa^n(T) \cup \{\varphi\}) \end{cases}$$

where $n = \sup\{k \in \mathbb{N} \mid Mod(\kappa^k(T) \cup \{\varphi\}) \neq Triv\}$;

$$\varphi \rhd_{\mathcal{C}_{lnr}} \psi \iff \begin{cases} Mod(T \cup \{\psi\}) \neq Triv, \text{ and} \\ Mod(T \cup \{\psi\}) \subseteq Mod(\kappa^m(T \cup \{\varphi\})) \end{cases}$$

where $m = \sup\{k \in \mathbb{N} \mid Mod(\kappa^k(T \cup \{\varphi\})) \neq Triv\}$.

Corollary 2. *The explanation relation $\rhd_{\mathcal{C}_{lcr}}$ satisfies all the logical properties of Theorems 2, 3 and 4.*

Proof. Again this is derived from the fact that for every $\varphi \in Sen$, $\mathcal{C}^{\varphi}_{lcr}$ has a unique maximal chain. It is easy to show that for every $k \in \mathbb{N}$, we have that $Mod(\kappa^k(\bigwedge T) \wedge \varphi') = Mod(\kappa^k(\bigwedge T) \wedge \varphi) \cap Mod(\varphi')$ when $Mod(\varphi') \subseteq Mod(\varphi)$ what proves Condition SUB-CUT of Theorem 4. Moreover, by definition of $\mathcal{C}_{lcr}$, we have trivially that for every $\varphi \in Sen$, all $\mathbb{M} \in \mathcal{C}_{\varphi}$ are finitely axiomatizable. Indeed, all of them are defined as $Mod(\kappa^k(T) \wedge \varphi)$ for $k \in \mathbb{N}$. Finally, only Property (8) of Theorem 3 requires for $\preceq_{\mathcal{C}}$ to be total. This has been proved for $\preceq_{\mathcal{C}^{\varphi}_{lcr}}$ in Proposition 3. □

Corollary 3. *The explanation relation $\rhd_{\mathcal{C}_{lnr}}$ satisfies all the logical properties of Theorems 2 and 3.*

As for $\rhd_{\mathcal{C}_{lcr}}$, the main reason for this result is that $\mathcal{C}^{\varphi}_{lnr}$ has a unique maximal chain and the associated ordering relation is total.

An example showing that these two definitions may provide different explanations of the same φ is given in the case of **PL** in Section 5.1.

Let us summarize the results in Table 1. Note that the properties of the two proposed explanatory relations $\rhd_{\mathcal{C}_{lcr}}$ and $\rhd_{\mathcal{C}_{lnr}}$ are in accordance with our first results established for the propositional logic **PL** in [9]. Counter-examples for the properties that do not hold for $\rhd_{\mathcal{C}_{lnr}}$ can be found in [9].

5. Applications

In this section, we illustrate our general approach by defining abduction based on retractions for the logics **PL**, **HCL**, **FOL**, **MPL** and the **DL** $\mathcal{ALC}$.

Table 1
Links between rationality postulates in [35] and properties in Theorems 2–4, and properties satisfied by $\rhd_{c_{lcr}}$ and $\rhd_{c_{lnr}}$.

Rationality postulates	Properties in Th. 2–4	$\rhd_{c_{lcr}}$	$\rhd_{c_{lnr}}$
LLE and RLE	(1)	√	√
RS	(2)	√	√
E-Con	(3)	√	√
E-Con	(4)	√	√
∼ RS	(5)	√	√
	(6)	√	√
∼ RS	(7)	√	√
ROR	(8)	√	√
E-Reflexivity	(9)	√	√
E-CM	(10)	√	
E-C-Cut	(11)	√	
	(12)	√	
	(13)	√	

5.1. Explanatory relations based on retraction in **PL**

Here, drawing inspiration from Bloch & al.'s work in [8–10] on Morpho-Logics, we define retractions based on erosions from mathematical morphology [7]. To define retractions in **PL**, we will apply set-theoretic morphological operations. First, let us recall basic definitions of erosion in mathematical morphology [7]. In complete lattices, an algebraic erosion is an operator that commutes with the infimum of the lattices. Concrete definitions of erosions often involve the notion of structuring element. Let us first consider the case of a lattice defined as the power set of some set (e.g. $\mathbb{R}^n$), with the inclusion relation. Let X and B be two subsets of $\mathbb{R}^n$. The erosion of X by the structuring element B, denoted by $E_B(X)$, is defined as follows:

$$E_B(X) = \{x \in \mathbb{R}^n \mid B_x \subseteq X\}$$

where B_x denotes the translation of B at x. More generally, erosions in any space can be defined in a similar way by considering the structuring element as a binary relationship between elements of this space.

In **PL**, knowing that we can identify any propositional formula φ with its set of interpretations $Mod(\varphi)$, this leads to the following erosion of a formula φ:

$$Mod(E_B(\varphi)) = \{\nu \in Mod \mid B_\nu \subseteq Mod(\varphi)\}$$

where B_ν contains all the models that satisfy some relationship with ν. The relationship standardly used is based on a discrete distance δ between models, and the most commonly used is the Hamming distance d_H where $d_H(\nu, \nu')$ for two propositional models over a same signature Σ is the number of propositional symbols that are instantiated differently in ν and ν'. In this case, we can rewrite the erosion of a formula as follows:

$$Mod(E_B(\varphi)) = \{\nu \in Mod \mid \forall \nu' \in Mod, \delta(\nu, \nu') \leq 1 \Rightarrow \nu' \in Mod(\varphi)\}$$

This consists in using the distance ball of radius 1 as structuring element. To ensure the non-consistency condition to our retraction based on erosion, we need to add a condition on distances, the *differentiation property*.

Definition 11 *(Differentiation property).* Let δ be a discrete distance over a set S. Let us note Γ_x for every $x \in S$, the set $\Gamma_x = \{y \in S \mid \delta(x, y) \leq 1\}$. The distance δ has the **differentiation property** if for every $x, y \in S$ such that $x \neq y$, $\Gamma_x \neq \Gamma_y$.

The Hamming distance trivially satisfies the differentiation property if the language has at least two propositional variables.

Proposition 4. *E_B is a retraction for finite signatures Σ, and when it is based on a distance between models that satisfies the differentiation property.*

Proof. It is anti-extensive since any erosion defined from a reflexive relationship is anti-extensive. Since δ is a distance, $\delta(\nu, \nu) = 0$ for any ν and $\nu \in B_\nu$, and thus for every φ and for every model $\nu \in Mod(E_B(\varphi))$, we have that $\nu \in Mod(\varphi)$.

Let φ be a propositional formula such that $Mod(\varphi) \neq Mod(\Sigma)$. As δ satisfies the differentiation property, there necessarily exists a model $\nu \in Mod(\varphi)$ and a model $\nu' \in Mod(\Sigma) \setminus Mod(\varphi)$ such that $\delta(\nu, \nu') \leq 1$ and $\nu' \notin Mod(\varphi)$. Hence, each

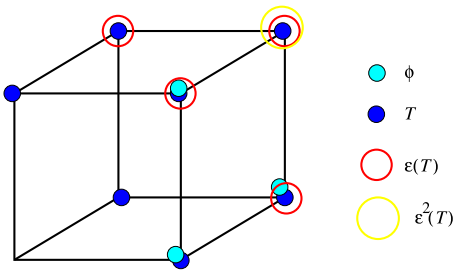


Fig. 2. An example of last consistent erosion.

application of E_B removes at least one model. As Σ is a finite signature, $Mod(\varphi)$ is finite, and therefore there is $k \in \mathbb{N}$ such that $Mod(E_B^k(\varphi)) = \emptyset$.[6] □

Let us first illustrate the instantiation of the two proposed definitions of explanation, when the retraction is an erosion using a ball of the Hamming distance as structuring element. Let us consider three propositional variables a, b, c, a KB $T = \{a \vee b \vee c\}$, and the observation to be explained $\varphi = (a \wedge \neg b \wedge c) \vee (a \wedge b \wedge \neg c) \vee (a \wedge \neg b \wedge \neg c)$. Models can be graphically represented as the vertices of a cube, as shown in Fig. 2 (for instance the bottom left vertex is $\neg a \wedge \neg b \wedge \neg c$ while the top right one is $a \wedge b \wedge c$).

It is easy to show that $\varepsilon(T)$ is consistent with φ, but $\varepsilon^2(T)$ is not [9]. Hence, for $\rhd_{\mathcal{C}_{lcr}}$ explanations ψ are such that $Mod(\psi) \subseteq \{(a \wedge \neg b \wedge c) \vee (a \wedge b \wedge \neg c)\}$. Similarly, it is easy to see that $\varepsilon(T \wedge \varphi) = \bot$, hence the explanations ψ for $\rhd_{\mathcal{C}_{lnr}}$ are such that $Mod(\psi) \subseteq Mod(T \wedge \varphi)$. In particular $\psi = (a \wedge \neg b \wedge \neg c)$ is a potential explanation of φ for $\rhd_{\mathcal{C}_{lnr}}$ but not for $\rhd_{\mathcal{C}_{lcr}}$. This is an example where the two proposed explanatory relations introduced in Definition 10 provide different results.

Let us now show that the choice of the structuring element used in the erosions can impact the obtained explanations. This example is adapted from [9]. Let us consider the explanatory relation $\rhd_{\mathcal{C}_{lcr}}$, the KB $T = \{a \Rightarrow c, b \Rightarrow c, a \vee b\}$, and the observation c.

(1) With the standard ball $B_\omega = \{\omega' \in Mod \mid d_H(\omega, \omega') \leq 1\}$, where d_H denotes the Hamming distance, we get $\varepsilon^1(T) = \bot$. Thus, we have in particular,

$$c \rhd_{\mathcal{C}_{lcr}} (a \vee b).$$

(2) Now we use $B_\omega^{ab} = \{\omega' \in B_\omega \mid \omega(x) = \omega'(x) \text{ for all } x \notin \{a, b\}\}$, i.e. B_ω^{ab} contains the valuations in B_ω that agree with ω outside $\{a, b\}$. Then $\varepsilon^1(T) = a \wedge b \wedge c$ and $\varepsilon^2(T) = \bot$. Thus

$$c \rhd_{\mathcal{C}_{lcr}} (a \wedge b).$$

Notice that $c \not\rhd_{\mathcal{C}_{lcr}} (a \vee b)$.

(3) Finally, let us consider the following structuring element

$$B_{\omega,2}^{ab} = \{\omega\} \cup \{\omega' \in Mod \mid d_H(\omega, \omega') = 2 \text{ and } \omega(x) = \omega'(x) \text{ for all } x \notin \{a, b\}\}$$

Then $\varepsilon^1(T) = \varepsilon^2(T) = (\neg a \wedge b \wedge c) \vee (a \wedge \neg b \wedge c)$. Thus,

$$c \rhd_{\mathcal{C}_{lcr}} (a \wedge \neg b) \vee (\neg a \wedge b).$$

Notice that $c \not\rhd_{\mathcal{C}_{lcr}} (a \wedge b)$.

Since the erosion has a fixed point ($\varepsilon^1(T) = \varepsilon^2(T) \ldots = \varepsilon^n(T) \ldots$), it does not satisfy the vacuum property, hence it is not a retraction. This is an example of explanation, still using cuttings, but not defined from a retraction.

These different results can be interesting in situations where different explanations may be expected. This is illustrated by the following examples, that are further discussed in [9]:

(1)

$$\begin{aligned} a &= \textit{rained\_last\_night} \\ b &= \textit{sprinkle\_was\_on} \\ c &= \textit{grass\_is\_wet} \end{aligned}$$

The "common sense cautious explanation" of c is $a \vee b$.

[6] As the negation is considered in **PL**, the set $Triv$ is empty, and the consistency of a formula φ can be defined by the fact that $Mod(\varphi) = \emptyset$.

(2)

a = *low_taxes*
b = *investment_increases*
c = *economy_grows*

An explanation that enhances the chances of achieving the goal of making the economy to grow is $a \wedge b$.

(3)

a = *book_was_left_somewhere else*
b = *somebody_took_the book*
c = *book_is_not_in_the shelf*

An explanation based on the principle of the "Ockham's razor" will select either a or b but not both, that is to say, $(a \wedge \neg b) \vee (\neg a \wedge b)$.

5.2. Explanatory relations based on retraction in **HCL**

Following our work in [1] on the definition of revision operators based on relaxations for **HCL**, we propose here to extend retractions that we have defined in the framework **PL** to deal with the Horn fragment of propositional formulas. First, let us recall some useful notions.

Definition 12 *(Model intersection).* Given a propositional signature Σ and two Σ-models $\nu, \nu' : \Sigma \to \{0,1\}$, we note $\nu \cap \nu' : \Sigma \to \{0,1\}$ the Σ-model defined by:

$$p \mapsto \begin{cases} 1 & \text{if } \nu(p) = \nu'(p) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Given a set of Σ-models $\mathcal{S}$, we note

$$cl_{\cap}(\mathcal{S}) = \mathcal{S} \cup \{\nu \cap \nu' \mid \nu, \nu' \in \mathcal{S}\}$$

which is the closure of $\mathcal{S}$ under intersection of positive atoms.

For any set $\mathcal{S}$ closed under intersection of positive atoms, there exists a Horn sentence φ that defines $\mathcal{S}$ (i.e. $Mod(\varphi) = \mathcal{S}$). Given a distance δ between models, we then define a retraction κ as follows: for every Horn formula φ, $\kappa(\varphi)$ is any Horn formula φ' such that $Mod(\varphi') = cl_{\cap}(Mod(E_B(\varphi))$, where E_B is the erosion defined as in **PL**, for B being a ball of δ, (by the previous property, we know that such a formula φ' exists).

Proposition 5. *With the same conditions as in Proposition 4, the mapping ρ is a retraction.*

Again both explanatory relations $\rhd_{lcr}$ and $\rhd_{lnr}$ can be defined from κ using Definition 10.

5.3. Explanatory relations based on retraction in **FOL**

A trivial way to define a retraction in **FOL** is to map any formula to an antilogy. A less trivial and more interesting retraction consists in replacing existential quantifiers involved in the formula to be retracted by universal ones. A dual approach has been adopted in [1] for defining revision operators using dilations in **FOL**. In the following we suppose that, given a signature, every formula φ in *Sen* is either a conjunction or a disjunction of formulas in prenex form (i.e. φ is either of the form $\bigwedge_j Q_1^j x_1^j \dots Q_{n_j}^j x_{n_j}^j . \psi_j$ or $\bigvee_j Q_1^j x_1^j \dots Q_{n_j}^j x_{n_j}^j . \psi_j$ where each Q_i^j is in $\{\forall, \exists\}$). Let us define the retraction κ as follows, for an antilogy τ:

- $\kappa(\tau) = \tau$;
- $\kappa(\forall x_1 \dots \forall x_n . \varphi) = \tau$;
- Let $\varphi = Q_1 x_1 \dots Q_n x_n . \psi$ be a formula such that the set $E_\varphi = \{i, 1 \le i \le n \mid Q_i = \exists\} \neq \emptyset$. Then, $\kappa(Q_1 x_1 \dots Q_n x_n . \varphi) = \bigvee_{i \in E_\varphi} \varphi_i$ where $\varphi_i = Q'_1 x_1 \dots Q'_n x_n . \psi$ such that for every $j \neq i$, $1 \le j \le n$, $Q'_j = Q_j$ and $Q'_i = \forall$;
- $\kappa(\bigwedge_j Q_1^j x_1^j \dots Q_{n_j}^j x_{n_j}^j . \psi) = \bigwedge_j \kappa(Q_1^j x_1^j \dots Q_{n_j}^j x_{n_j}^j . \psi)$.
- $\kappa(\bigvee_j Q_1^j x_1^j \dots Q_{n_j}^j x_{n_j}^j . \psi) = \bigvee_j \kappa(Q_1^j x_1^j \dots Q_{n_j}^j x_{n_j}^j . \psi)$.

Proposition 6. *κ is a retraction.*

Proof. κ is obviously anti-extensive, and satisfies the vacuum property because in a finite number of steps, we always reach the antilogy τ. □

Example 5. To illustrate our approach, let us consider the example taken from [14] and defined by the KB which only contains the formula $\forall x.\forall y.\forall z.(p(x,y) \wedge p(y,z) \Rightarrow p(x,z))$ and the observation $\varphi = \exists w.p(w,w)$. According to the explanatory relation we consider (i.e. either $\rhd_{lcr}$ or $\rhd_{lnr}$), the retracted formula will be different. For $\rhd_{lcr}$, only the formula $\forall x.\forall y.\forall z.(p(x,y) \wedge p(y,z) \Rightarrow p(x,z))$ is retracted. But in this case, to preserve consistency, the maximum number of retraction steps to apply is 0. Hence, we have many possible explanations such as the trivial one $\exists w.p(w,w)$ (i.e. $\varphi \rhd_{lcr} \varphi$). The minimal explanation $\exists x.\exists y.p(x,y) \wedge p(y,x)$ given in [14] also satisfies $\varphi \rhd_{lcr} \exists x.\exists y.p(x,y) \wedge p(y,x)$.

For $\rhd_{lnr}$, we can directly retract the formula $\forall x.\forall y.\forall z.(p(x,y) \wedge p(y,z) \Rightarrow p(x,z)) \wedge \exists w.p(w,w)$, but in this case to preserve consistency, the maximum number of retraction steps is 0, and we come up with the previous case. Now, we can also consider the prenex form of $\forall x.\forall y.\forall z.(p(x,y) \wedge p(y,z) \Rightarrow p(x,z)) \wedge \exists w.p(w,w)$ which is $\forall x.\forall y.\forall z.\exists w.(p(x,y) \wedge p(y,z) \Rightarrow p(x,z)) \wedge p(w,w)$. Here, to preserve consistency, the maximum number of retraction steps to apply is 1. We then obtain the formula $\forall x.\forall y.\forall z.\forall w.(p(x,y) \wedge p(y,z) \Rightarrow p(x,z)) \wedge p(w,w)$, and a possible explanation here is $\varphi \rhd_{lnr} \forall w.p(w,w)$. In contrast, we now have that $\varphi \not\rhd_{lnr} \exists w.p(w,w)$ and $\varphi \not\rhd_{lnr} \exists x.\exists y.p(x,y) \wedge p(y,x)$.

5.4. Explanatory relations based on retraction in **MPL**

By the classical first-order correspondence of **MPL**, we can easily adapt the retraction defined for **FOL** by replacing $\Diamond$ by $\Box$. Now, we can go further when dealing with formulas of the form $\Box \ldots \Box\varphi$. Indeed, in **MPL**, we have that $Mod(\varphi) \subseteq Mod(\Box\varphi)$. Hence, we can remove in formulas the most external $\Box$. Of course, when dealing with modal logics such as T, $S4$, B and $S5$ (i.e. the accessibility relation of Kripke models is always reflexive) where the formula $\Box\varphi \Rightarrow \varphi$ is a tautology, this is of no interest because in this case we have that $\varphi \equiv \Box\varphi$. This gives rise to the following retraction κ: here also we suppose that every formula φ in Sen is either a conjunction or a disjunction of formulas φ in the following normal form $\varphi = M_1 \ldots M_n.\psi$ where each M_i is in $\{\Box, \Diamond\}$).

- $\kappa(\tau) = \tau$ if τ is any antilogy;
- $\kappa(\varphi) = \tau$ if φ is modality free;
- $\kappa(\Box\varphi) = \varphi$;
- Let $\varphi = M_1 \ldots M_n\psi$ be a formula such that the set $E_\varphi = \{i, 1 \leq i \leq n \mid M_i = \Diamond\} \neq \emptyset$. Then, $\kappa(M_1 \ldots M_n\varphi) = \bigvee_{i \in E_\varphi} \varphi_i$ where $\varphi_i = M'_1 \ldots M'_n.\psi_i$ such that for every $j \neq i$, $1 \leq j \leq n$, $M'_j = M_j$ and $M'_i = \Box$;
- $\kappa(\bigwedge_j M_1^j \ldots M_{n_j}^j \psi) = \bigwedge_j \kappa(M_1^j \ldots M_{n_j}^j \psi)$.
- $\kappa(\bigvee_j M_1^j \ldots M_{n_j}^j \psi) = \bigvee_j \kappa(M_1^j \ldots M_{n_j}^j \psi)$.

Proposition 7. *κ is a retraction.*

Proof. The proof is similar to the one of Proposition 6 and relies on the fact that $Mod(\varphi) \subseteq Mod(\Box\varphi)$, κ is anti-extensive and satisfies the vacuum property. □

5.5. Explanatory relations based on retraction in **DL**

Abduction in DL can take different forms: concept abduction, TBox abduction, ABox abduction and knowledge base abduction (see e.g. [23,29,32]).

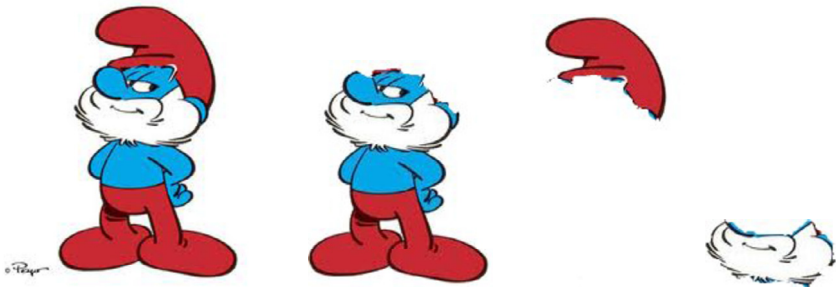
Definition 13 *(Abduction types in DL).* Let $\mathcal{L}, \mathcal{L}'$ be two arbitrary description logics, $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ a knowledge base in $\mathcal{L}$ with $\mathcal{T}$ the TBox and $\mathcal{A}$ the Abox, C, D two concepts in $\mathcal{L}$ satisfiable with respect to $\mathcal{K}$ (i.e. that admit non empty interpretations). Abduction forms in DL are as follows:
– *Concept abduction:* given an observation concept O in $\mathcal{L}$ satisfiable w.r.t. $\mathcal{K}$, the set of explanations introduced in Definition 4 writes as:

$$Expla_{\mathcal{K}}(O) = \{H \mid H \text{ satisfiable w.r.t. } \mathcal{K} \text{ and } \mathcal{K} \models H \sqsubseteq O\}.$$

The set of concepts in $Expla_{\mathcal{K}}(O)$ may possibly be expressed in another description logic $\mathcal{L}'$.
– *TBox abduction:* let $C \sqsubseteq D$ be satisfiable w.r.t. $\mathcal{K}$, the set of explanations is made of axioms defined as:

$$Expla_{\mathcal{K}}(C \sqsubseteq D) = \{E \sqsubseteq F \mid E \sqsubseteq F \text{ satisfiable w.r.t. } \mathcal{K} \text{ and } \mathcal{K} \cup \{E \sqsubseteq F\} \models C \sqsubseteq D\}.$$

– *ABox abduction:* let S_a be a set of assertions representing the observation, the set of explanations is the set S_b of ABox assertions such that S_b satisfiable w.r.t. $\mathcal{K}$ and $\mathcal{K} \cup S_b \models S_a$.



(a) Complete picture (b) Region a (c) Region b (d) Region c

Fig. 3. Picture of Smurf and three regions a, b, c.

– *KB abduction:* let $\{\varphi\}$ be a consistent set of ABox or TBox assertions w.r.t. $\mathcal{K}$. A solution of knowledge base abduction, considered as a combination of TBox abduction and ABox abduction, is any finite set $S = \{\psi_i, i = 1...n\}$ in $\mathcal{L}'$ satisfiable w.r.t. $\mathcal{K}$ and such that $\mathcal{K} \cup S \models \{\varphi\}$.

As in any other logic, additional constraints can be used to find the preferred explanations in $Expla_{\mathcal{K}}$ (minimality, etc. (see e.g. [5]).

Let us illustrate these notions on an example inspired from an image interpretation task.

Example 6. [7] Suppose we have an image, in which we have identified three regions a, b and c (Fig. 3). Region b has been identified as `Hat` and has a color attribute `Red`, while region c has been identified as `Beard`. There is a spatial relation `hasOnTop` that links a to b, and a spatial relation `hasPart` linking a to c. Furthermore, the background knowledge tells us that smurf leaders are smurfs that wear red hats (i.e. have them on top) and have beards. In this example, a good approach should be able to come up with the explanation that region a might be a smurf leader.

The background knowledge is encoded as a TBox:

$$\mathcal{T} = \{\texttt{SmurfLeader} \sqsubseteq \exists \texttt{hasPart}.\texttt{Beard} \sqcap \exists \texttt{hasOnTop}.\texttt{RedHat},$$
$$\texttt{RedHat} \equiv \texttt{Hat} \sqcap \exists \texttt{hasColor}.\texttt{Red}\}$$

The observation is encoded as an ABox

$$\mathcal{A}_o = \{(a, b) : \texttt{hasOnTop},$$
$$(a, c) : \texttt{hasPart},$$
$$b : \texttt{Hat},$$
$$b : \exists \texttt{hasColor}.\texttt{Red},$$
$$c : \texttt{Beard}\}$$

A preferred explanation for this example is $\mathcal{A}_e = \{a : \texttt{SmurfLeader}\}$.

In Concept abduction, we need to say which region we want to explain, here region a. Then the *most specific concept* for this region is:

$$O = msc(a) = \exists hasPart.Beard \sqcap \exists hasOnTop.(Hat \sqcap hasColor.Red).$$

Concept Abduction now looks for a concept description C such that $\mathcal{T} \models C \sqsubseteq O$. A solution that is both length minimal[8] and $\sqsubseteq$-maximal[9] would be $C = SmurfLeader$ which is one of the expected solution.

In this paper, we consider the general form of abduction in DL, i.e. KB abduction. The other forms can be seen as particular cases. Explanatory relations of a KB in DL can be defined in two ways:

- When the logic is equipped with the disjunction and full negation constructors, as it is the case of the logic $\mathcal{ALC}$ and its extensions, the theory is transformed into an *internalized concept* on which the retraction operators act. The internalized

[7] This example results from a discussion with Felix Distel during his visit at LTCI, Télécom ParisTech (summer 2013).

[8] The length of a concept is defined as the number of atomic concepts appearing in it.

[9] An explanation H is $\sqsubseteq$-maximal if there is no other explanation H' such that $H \sqsubseteq H'$.

concept is defined as follows: $C_{\mathcal{T}} := \bigsqcap_{(C \sqsubseteq D) \in \mathcal{T}} (\neg C \sqcup D)$. When Abox assertions are considered one can also internalize the Abox to an equivalent concept provided that nominals are part of the syntax. Nominals are concept descriptions having as semantics: $(\{o\})^{\mathcal{I}} = \{o^{\mathcal{I}}\}$, where $(\_^{\mathcal{I}}, \Delta^{\mathcal{I}})$ is an interpretation. Then the Abox assertions are transformed into concept inclusions as follows: $a : C$ corresponds to $\{a\} \sqsubseteq C$, $(a, b) : r$ corresponds to $\{a\} \sqsubseteq \exists r.\{b\}$, etc.

- When the logic does not allow for full negation, a possible workaround consists in retracting all the formulas at the same time. Hence, one needs to define concrete retraction operators both on concepts and on formulas. Σ-formula retraction can be defined in two ways (other definitions may also exist). For sentences of the form $C \sqsubseteq D$, a first possible approach consists in retracting the set of models of D while the second one amounts to "relax" the set of models of C (see e.g. [1] for definitions of relaxations in satisfaction systems, with several examples in DL).

Note also that retracting a concept (or formula) amounts to "relaxing" its negated form, and can then be seen as its dual operator. The notion of concept relaxation has been first introduced in description logics to define dissimilarity measures between concepts in [18,19], and has been extended to define revision operators in arbitrary logics in [1]. It is extensive and exhaustive, i.e. $\exists k \in \mathbb{N}$ such that $\rho^k(C) \equiv \top$, where ρ^k denotes k iterations of a relaxation ρ. Relaxation of formulas have been defined from retraction of concepts to come up with revision operators, in particular within the context of description logics. In [1], some retraction operators of DL-concept descriptions have been introduced. These operators, designed for the purpose of revision are too strong, since their aim is to remove formulas in the background knowledge that are inconsistent with the new acquired one. The philosophy behind abduction is slightly different. One should add new knowledge to the set of consequences of the background theory. Hence, the retraction operator should act on the entire KB rather than just seeking a subpart that is more appropriate to revise. Concretely, this means that instead of relaxing the formulas, potentially each one to a different extent, as done for revision, for abduction, all formulas have to be retracted in the same way.

Hence, we need to introduce new retraction operators suited to the purpose of abduction. In what follows, we restrict ourselves to the context of the logic $\mathcal{ALC}$, as defined in Section 2, possibly enriched with nominals.

Definition 14 *(Concept retraction).* Let $\mathsf{C}(\Sigma)$ be the set of concept descriptions defined over a signature Σ. A *(concept) retraction* is an operator $\kappa : \mathsf{C}(\Sigma) \to \mathsf{C}(\Sigma)$ that satisfies the following two properties for all $C \in \mathsf{C}(\Sigma)$ such that C is not equivalent to $\top$:

(1) κ is *anti-extensive*, i.e. $\kappa(C) \sqsubseteq C$, and
(2) κ satisfies the *vacuum* property, i.e. $\exists k \in \mathbb{N}, \kappa^k(C) = \bot$ where κ^0 is the identity mapping, and for all $k > 0$, $\kappa^k(C) = \kappa(\kappa^{k-1}(C))$.

This definition is a direct instantiation to DL of Definition 9. Now we propose, as an example, the following operator to define a particular retraction in $\mathcal{ALC}$.

Definition 15. Given an $\mathcal{ALC}$-concept description C we define an operator κ_f recursively as follows.

- For $C = A \in N_C$ (i.e. an atomic concept), $\kappa_f(C) = \bot$.
- For $C = \neg A$, $\kappa_f(C) = \bot$.
- For $C = \bot$, $\kappa_f(C) = \bot$.
- For $C = \top$ $\kappa_f(C) = \top$.
- For $C = C_1 \sqcup C_2$, $\kappa_f(C_1 \sqcup C_2) = \big(\kappa_f(C_1) \sqcup C_2\big) \sqcap \big(C_1 \sqcup \kappa_f(C_2)\big)$.
- For $C = C_1 \sqcap C_2$, $\kappa_f(C_1 \sqcap C_2) = \kappa_f(C_1) \sqcap \kappa_f(C_2)$.
- For $C = \forall r.D$, with $r \in N_R$, $\kappa_f(C) = \forall r.\kappa_f(D)$.
- For $C = \exists r.D$, $\kappa_f(C) = (\forall r.D) \sqcup \big(\exists r.\kappa_f(D)\big)$.

Note that this definition assumes that any concept is rewritten using standard De Morgan rules so that negations apply only on atomic concepts.

Proposition 8. *The operator κ_f is a retraction.*

Proof. The proof is straightforward, by induction on the structure of C. □

Example 7. Let us illustrate this retraction operator on Example 6 using the explanatory relations introduced in Section 4. To ease the reading, we will note the concepts by capital letters and roles by small letters, e.g. B=`Beard`, S=`SmurfLeader`, H=`Hat`, R=`Red`, t=`hasOnTop`, p=`hasPart`, c=`hasColor`. The unfolded TBox writes as:

$$\mathcal{T} = \{S \sqsubseteq \exists p.B \sqcap \exists t.(H \sqcap \exists c.R)\}$$

and the observation writes as:

$$\varphi = \exists p.B \sqcap \exists t.(H \sqcap \exists c.R)$$

Note that $\mathcal{T} = \{S \sqsubseteq \varphi\}$, and $C_{\mathcal{T}} = \neg S \sqcup \varphi$, where $C_{\mathcal{T}}$ is the internalized concept of the TBox.

First, note that we have:

$$\kappa_f(\varphi) = \kappa_f(\exists p.B \sqcap \exists t.(H \sqcap \exists c.R)) = \forall p.B \sqcap \forall t.(H \sqcap \exists c.R) \tag{5}$$

and

$$\kappa_f^2(\varphi) = \kappa_f(\kappa_f(\varphi)) = \kappa_f(\forall p.B \sqcap \forall t.(H \sqcap \exists c.R)) = \bot \tag{6}$$

Let us now consider the two explanatory relations $\rhd_{\mathcal{C}_{lcr}}$ and $\rhd_{\mathcal{C}_{lnr}}$ defined in Section 4, applied here in $\mathcal{ALC}$ and with κ_f.

Case 1 ($\rhd_{\mathcal{C}_{lcr}}$). The $\rhd_{\mathcal{C}_{lcr}}$ relation amounts to take the last retraction of $C_{\mathcal{T}}$ that is still consistent with φ, i.e.:

$$\varphi \rhd_{\mathcal{C}_{lcr}} \psi \Leftrightarrow \psi \sqsubseteq \kappa_f^n(\neg S \sqcup \varphi) \sqcap \varphi$$

We have:

$$\begin{aligned}
\kappa_f^1(\neg S \sqcup \varphi) &= \varphi \sqcap (\neg S \sqcup \kappa_f^1(\varphi)) \\
\kappa_f^2(\neg S \sqcup \varphi) &= \kappa_f^1(\varphi) \sqcap (\kappa_f^1(\varphi) \sqcap (\kappa_f^2(\varphi) \sqcup \neg S)) \\
&= \kappa_f^1(\varphi) \sqcap (\kappa_f^2(\varphi) \sqcup \neg S)) \\
&= \kappa_f^1(\varphi) \sqcap \neg S, \text{ since } \kappa_f^2(\varphi) = \bot \\
\kappa_f^3(\neg S \sqcup \varphi) &= \bot
\end{aligned}$$

Then $n = 2$ and

$$\psi \sqsubseteq (\kappa_f^1(\varphi) \sqcap \neg S) \sqcap \varphi$$

with $\kappa_f^1(\varphi) = \forall p.B \sqcap \forall t.(H \sqcap \exists c.R)$. Since κ_f is anti-extensive, $\kappa_f^1(\varphi) \sqcap \varphi = \kappa_f^1(\varphi)$, and

$$\psi \sqsubseteq (\forall p.B \sqcap \forall t.(H \sqcap \exists c.R) \sqcap \neg S)$$

Case 2 ($\rhd_{\mathcal{C}_{lnr}}$). The $\rhd_{\mathcal{C}_{lnr}}$ relation amounts to take the last non empty retraction of $C_{\mathcal{T}} \sqcap \varphi$, i.e.:

$$\varphi \rhd_{\mathcal{C}_{lnr}} \psi \Leftrightarrow \psi \sqsubseteq \kappa_f^n((\neg S \sqcup \varphi) \sqcap \varphi)$$

with the largest possible value of n such that the retraction is not empty. Since we have $(\neg S \sqcup \varphi) \sqcap \varphi = \varphi$, according to Equations (5) and (6), $n = 1$, and therefore

$$\psi \sqsubseteq \forall p.B \sqcap \forall t.(H \sqcap \exists c.R) \sqcup (S \sqcap \neg\varphi)$$

A possible solution is then S according to subset minimality, which well fits the intuition.

Several other retractions could be proposed. In particular, several relaxations proposed in [1] for revision could be modified to become retractions. For instance, relaxing $C \sqsubseteq D$ can be performed either by retracting C or by relaxing D. Similarly, retracting $C \sqsubseteq D$ could be performed either by relaxing C or by retracting D.

6. Conclusion

In this paper, we proposed a new framework for abduction in satisfaction systems, by introducing the notion of cutting, which provides a structure on the set of models among which an explanation can be found. Inspired by previous work in propositional logic where abduction was defined from morphological erosions, we proposed to define cuttings from the more general notion of retraction, and proved a set of rationality postulates for the derived explanatory relations. The generic feature of the proposed approach has been illustrated by providing concrete examples of retractions, cuttings and explanatory relations in various logics.

Future work will aim at further analyzing the structure of the set of cuttings $\mathcal{C}$ for a theory T, and the properties of the derived relation $\preceq_{\mathcal{C}}$. The examples in DL could also be further investigated, by considering other types of retractions as well as various fragments of $\mathcal{ALC}$, as done for revision in [1]. Links between the proposed approach with other abduction

methods could also deserve to be investigated, such as with sequent calculus, prime implicants [38] or equational logic [20, 21]. To address the question of uncertainty in the observations, or in the theory, the proposed approach could be extended to the case of fuzzy logic, based on our previous work on mathematical morphology in the framework of institutions [2], or alternatively to the case of defeasible DL [13]. Finally applications will be further developed, in particular for image understanding and spatial reasoning.

Acknowledgements

This work was partially funded by a grant from Labex DIGICOSME (ANR-11-IDEX-0003-02). The fourth author thanks the LTCI and Télécom ParisTech who have given partial funding during several sabbatical visits, and the CDCHT-ULA who has given partial funding through the Project C-1855-13-05-A.

References

[1] M. Aiguier, J. Atif, I. Bloch, C. Hudelot, Belief revision, minimal change and relaxation: a general framework based on satisfaction systems, and applications to description logics, Artif. Intell. 256 (2018) 160–180.
[2] M. Aiguier, I. Bloch, Dual logic concepts based on mathematical morphology in stratified institutions: applications to spatial reasoning, Tech. Rep., arXiv:1710.05661, Oct. 2017, CoRR.
[3] A. Aliseda-Liera, Abduction in Logic, Philosophy of Science and Artificial Intelligence, Ph.D. thesis, Stanford University, 1997.
[4] J. Atif, C. Hudelot, I. Bloch, Explanatory reasoning for image understanding using formal concept analysis and description logics, IEEE Trans. Syst. Man Cybern. Syst. 44 (5) (2014) 552–570.
[5] M. Bienvenu, Complexity of abduction in the el family of lightweight description logics, in: International Conference on Principles of Knowledge Representation and Reasoning (KR), 2008, pp. 220–230.
[6] M. Bienvenu, Prime implicates and prime implicants from propositional to modal logic, Artif. Intell. Res. 36 (2009) 71–128.
[7] I. Bloch, H. Heijmans, C. Ronse, Mathematical morphology, in: M. Aiello, I. Pratt-Hartman, J. van Benthem (Eds.), Handbook of Spatial Logics, Springer-Verlag, 2006, pp. 857–947.
[8] I. Bloch, J. Lang, Towards mathematical morpho-logics, in: B. Bouchon-Meunier, J. Gutierrez-Rios, L. Magdalena, R. Yager (Eds.), Technologies for Constructing Intelligent Systems, Springer-Verlag, 2002, pp. 367–380.
[9] I. Bloch, J. Lang, R. Pino Pérez, C. Uzcátegui, Morphologic for knowledge dynamics: revision, fusion, abduction, Tech. Rep., arXiv:1802.05142v1 [cs.AI], Feb. 2018.
[10] I. Bloch, R. Pino Pérez, C. Uzcategui, A unified treatment of knowledge dynamics, in: International Conference on Principles of Knowledge Representation and Reasoning (KR), AAAI Press, 2004, pp. 329–337.
[11] R. Booth, D. Gabbay, S. Kaci, T. Rienstra, L.v.d. Torre, Abduction and dialogical proof in argumentation and logic programming, in: Twenty-First European Conference on Artificial Intelligence (ECAI), IOS Press, 2014, pp. 117–122.
[12] R. Booth, P. Mikolaj, Using distances for aggregation in abstract argumentation, in: 26th Benelux Conference on Artificial Intelligence (BNAIC), 2014.
[13] K. Britz, I.J. Varzinczak, Towards defeasible SROIQ, in: 30th International Workshop on Description Logics, 2017.
[14] M. Cialdea-Mayer, F. Pirri, First order abduction via tableau and sequent calculi, Bull. IGPL 1 (1) (1993) 99–117.
[15] M. Cialdea-Mayer, F. Pirri, Propositional abduction in modal logic, Log. J. IGPL 6 (3) (1995) 907–919.
[16] L. Console, P. Torasso, A spectrum of logical definitions of model-based diagnosis, Comput. Intell. 7 (3) (1991) 133–141.
[17] R. Diaconescu, Institution-Independent Model Theory, Universal Logic, Birkhäuser, 2008.
[18] F. Distel, J. Atif, I. Bloch, Concept dissimilarity on tree edit distance and morphological dilatations, in: European Conference on Artificial Intelligence (ECAI), 2014, pp. 249–254.
[19] F. Distel, J. Atif, I. Bloch, Concept dissimilarity with triangle inequality, in: C. Baral, G.D. Giacomo, T. Eiter (Eds.), Fourteenth International Conference on Principles of Knowledge Representation and Reasoning (KR), AAAI Press, 2014, pp. 614–617.
[20] M. Echenim, N. Peltier, A calculus for generating ground explanations, in: International Joint Conference on Automated Reasoning, Springer, 2012, pp. 194–209.
[21] M. Echenim, N. Peltier, S. Tourret, An approach to abductive reasoning in equational logic, in: International Joint Conference on Artificial Intelligence (IJCAI), 2013, pp. 531–537.
[22] T. Eiter, G. Gottlob, The complexity of logic-based abduction, J. ACM 42 (1) (1995) 3–42.
[23] C. Elsenbroich, O. Kutz, U. Sattler, A case for abductive reasoning over ontologies, in: OWL: Experiences and Directions, vol. 67, 2006, pp. 81–82.
[24] P.-A. Flach, Rationality postulates for induction, in: Y. Shoam (Ed.), Sixth Conference of Theoretical Aspects of Rationality and Knowledge, TARK-96, 1996, pp. 267–281.
[25] P.-A. Flach, Logical characterisations of inductive learning, in: D.-M. Gabbay, R. Kuse (Eds.), Abductive Reasoning and Learning, Kluwer Academic, 2000, pp. 155–196.
[26] P.-A. Flach, On the logic of hypothesis generation, in: P.-A. Flach, A. Kakas (Eds.), Abduction and Induction, Kluwer Academic, 2000, pp. 89–106.
[27] J.-A. Goguen, R.-M. Burstall, Institutions: abstract model theory for specification and programming, J. ACM 39 (1) (1992) 95–146.
[28] K. Halland, K. Britz, ABox abduction in ALC using a DL tableau, in: ACM South African Institute for Computer Scientists and Information Technologists Conference, 2012, pp. 51–58.
[29] K. Halland, K. Britz, S. Klarman, TBox abduction in $\mathcal{ALC}$ using a DL tableau, in: 27th International Workshop on Description Logics, DL-2014, 2014.
[30] S. Han, A. Hutter, W. Stechele, A reasoning approach to enable abductive semantic explanation upon collected observations for forensic visual surveillance, in: IEEE International Conference on Multimedia and Expo, 2011, pp. 1–7.
[31] J.R. Hobbs, Abduction in natural language understanding, in: Handbook of Pragmatics, 2004, pp. 724–741.
[32] S. Klarman, U. Endriss, S. Schlobach, ABox abduction in the description logic, J. Autom. Reason. 46 (1) (2011) 43–80.
[33] P. Marquis, Extending abduction from propositional to first-order logic, in: P. Jorrand, J. Kelemen (Eds.), Fundamentals of Artificial Intelligence Research, International Workshop (FAIR), in: LNCS, vol. 535, Springer-Verlag, 1991, pp. 141–155.
[34] C.S. Peirce, C. Hartshorne, P. Weiss, A.W. Burks, Collected Papers of Charles Sanders Peirce: Science and Philosophy, Belknap Press of Harvard University Press, 1958.
[35] R. Pino Pérez, C. Uzcátegui, Jumping to explanation versus jumping to conclusions, Artif. Intell. 111 (2) (1999) 131–169.
[36] R. Pino Pérez, C. Uzcátegui, Preferences and explanations, Artif. Intell. 149 (2003) 1–30.
[37] J. Pukancová, M. Homola, Tableau-based ABox abduction for the ALCHO description logic, in: 30th International Workshop on Description Logics, Montpellier, France, 2017.

[38] W.V. Quine, A way to simplify truth functions, Am. Math. Mon. 62 (9) (1955) 627–631.
[39] A.-L. Reyes-Cabello, A. Aliseda-Llera, A. Nepomuceno-Fernandez, Towards abductive reasoning in first-order logic, Log. J. IGPL 14 (2) (2006) 287–304.
[40] F. Soler-Toscano, A. Nepomuceno-Fernadez, A. Aliseda-Llera, Model-based abduction via dual resolution, Log. J. IGPL 14 (2) (2006) 305–319.
[41] Y. Yang, J. Atif, I. Bloch, Abductive reasoning using tableau methods for high-level image interpretation, in: German Conference on Artificial Intelligence, KI2015, Dresden, Germany, in: LNAI, vol. 9324, 2015, pp. 356–365.