Arguments using ontological and causal knowledge (FoIKS 2014)

Philippe Besnard, Marie-Odile Cordier, Yves Moinard

Arguments using
ontological and causal knowledge

Philippe Besnard   CNRS (UMR IRIT),
Université de Toulouse, France,
IRISA
besnard@irit.fr

Marie-Odile Cordier   Université de Rennes 1,
Inria,
IRISA
cordier@irisa.fr

Yves Moinard   Inria,
IRISA
yves.moinard@inria.fr

Address: IRISA, Campus de Beaulieu, 35042 Rennes, France

January 15, 2014

Abstract

We explore an approach to reasoning about causes via argumentation. We consider a causal model for a physical system, and we look for arguments about facts. Some arguments are meant to provide explanations of facts whereas some challenge these explanations and so on. At the root of argumentation here, are causal links ($\{A_1, \cdots, A_n\}$ *causes* B) and also ontological links (c_1 *is_a* c_2). We introduce here a logical approach which provides a candidate explanation ($\{A_1, \cdots, A_n\}$ *explains* $\{B_1, \cdots, B_m\}$) by resorting to an underlying causal link substantiated with appropriate ontological links. Argumentation is then at work from these various explanation links. A case study is developed: a severe storm Xynthia that devastated a county in France in 2010, with an unaccountably high number of casualties.

1   Introduction and Motivation

Looking for explanations is a frequent operation, in various domains, from judiciary to mechanical fields. We consider the case where we have a (not necessarily exhaustive) description of some mechanism, or situation, and we are looking for explanations of some facts. The description contains logical formulas, together with some *causal* and *ontological* formulas (or links). Indeed, it is well-known that, although there are similarities between *causation* and *implication*, causation cannot be rendered by a simple logical implication. Moreover, confusing causation and *co-occurrence* could lead to undesirable relationships. This is why we resort here to a *causal formalism* such that some *causal links* and *ontological links* are added to classical logical formulas. Then, our causal formalism will produce various *explanation links* [2].

Each causal link gives rise to explanation links, and each explanation link must appeal to at least one causal link. The ontology gives rise to further explanation links, although only in the case that these come from explanation links previously obtained: no explanation link can come only from ontological information. In fact, the ontology determines a new connective (it can be viewed as a strong implication) which can induce these further explanation links, whereas classical implication cannot. Indeed, given an explanation link, logical implication is not enough to derive from it other explanation links (apart from trivially equivalent ones).

Despite these restrictions, if the situation described is complex enough, there will be a large number of explanation links, i.e., possible explanations, and *argumentation* is an appealing approach to distinguish between all these explanations.

We introduce in section 2 an explicative model, built from a causal model and an ontological model. Section 4 shows how the explicative model produces explanations. Section 5 deals with argumentation about explanations and we conclude in section 6. Section 3 presents a case study: a severe storm called Xynthia, that resulted in 26 deaths in a single group of houses in La Faute sur Mer, a village in Vendée during a night in 2010.

2   Explicative Model = Causal Model + Ontological Model

The model that is used to obtain explanations and support argumentation, called the explicative model, is built from a causal model relating literals in causal links and from an ontological model where classes (which denote types of object as usual in the literature about ontology) are related by specialization/generalization links. Data consist of causal links and "is_a" relationships (specifying a hierarchy of classes, see the ontological model in section 2.3) and background knowledge (formulae of a sorted logic whose sorts are the classes of the aforementioned hierarchy).

From these data, tentative explanations are inferred according to principles using the so-called ontological deduction links obtained in the explicative model.

2.1 Closed literals

By a *closed literal*, we mean a propositional literal or a formula

$$\exists \, x : class \, \neg^{\{0,1\}} P(x) \qquad \text{or} \qquad \forall \, x : class \, \neg^{\{0,1\}} P(x)$$

where x is a variable and P is a unary predicate symbol, preceded or not by negation. Throughout, a closed literal of the form $\exists \, x : classP(x)$ is abbreviated as $\exists \, P(class)$ and $\forall \, x : classP(x)$ is abbreviated as $\forall \, P(class)$ and similarly for $\neg P$ instead of P. From now on, when we write simply *literal*, we mean a closed literal.

Lastly, extension to n-ary predicates is unproblematic except for heavy notation. Henceforth, it is not considered in this paper for the sake of clarity.

2.2 The causal model

By a causal model [11], we mean a representation of a body of causal relationships to be used to generate arguments that display explanations for a given set of facts. Intuitively, a causal link expresses that a bunch of facts causes some effect.

Notation 1 A *causal link* is of the form

$$\{\alpha_1, \alpha_2, \cdots, \alpha_n\} \, causes \, \beta$$

where $\alpha_1, \alpha_2, \cdots, \alpha_n, \beta$ are literals.

These causal links will be used in order to obtain *explanation links* in section 4.

Example 1 *My being a gourmet with a sweet tooth causes me to appreciate some cake can be represented by*

$$\{\texttt{sweet\_tooth\_gourmet}\} \quad causes \quad \exists \, X : \texttt{cake Ilike(X)}$$

Similarly, my being greedy causes me to appreciate all cakes can be represented by

$$\{\texttt{IamGreedy}\} \quad causes \quad \forall \, X : \texttt{cake Ilike(X)}$$

In the figures (e.g., part of the causal model for Xynthia in Fig. 2), each plain black arrow represents a causal link.

2.3 The ontological model

Our approach assumes an elementary ontology in which specialization/generalization links between classes are denoted $c_n \xrightarrow{\text{ISA}} c_m$.

Notation 2 An $\xrightarrow{\text{ISA}}$ link has the form $c_1 \xrightarrow{\text{ISA}} c_2$ where c_1 and c_2 are sorts in our logical language that denote classes such that c_1 is a subclass of c_2 in the ontology.

3

Thus, $\xrightarrow{\text{ISA}}$ denotes the usual specialization link between classes. E.g., we have `Hurri` $\xrightarrow{\text{ISA}}$ `SWind` and `House1FPA` $\xrightarrow{\text{ISA}}$ `HouseFPA` and `HouseFPA` $\xrightarrow{\text{ISA}}$ `BFPA`[1]: the class `Hurri` of hurricane is a specialization of the class `SWind` of strong wing, the class `House1FPA` of typical Vendée low houses with a single level in flood-prone area is a specialization of the class `HouseFPA` of houses in this area, which itself is a specialization of the class `BFPA` of buildings in this area.

In the figures (e.g., part of the ontological model for Xynthia in Fig. 1), each white-headed arrow labelled with *is-a* denotes an $\xrightarrow{\text{ISA}}$ link.

$$\text{The relation } \xrightarrow{\text{ISA}} \text{ is required to be transitive and reflexive.} \tag{1}$$

Reflexivity is due to technical reasons simplifying various definitions and properties.

2.4 The explicative model

The resulting model (causal model + ontological deduction link) is the *explicative model*, from which explanation links can be inferred.

Notation 3 An *ontological deduction link* has the form $\Phi_1 \xrightarrow{\text{DEDO}} \Phi_2$ where Φ_1 and Φ_2 are two sets of literals.

If $\Phi_1 = \{\varphi_1\}$ and $\Phi_2 = \{\varphi_2\}$ are singletons, we may actually omit curly brackets in the link $\{\varphi_1\} \xrightarrow{\text{DEDO}} \{\varphi_2\}$ to abbreviate it as $\varphi_1 \xrightarrow{\text{DEDO}} \varphi_2$.

Such a link between literals $\varphi_1 \xrightarrow{\text{DEDO}} \varphi_2$ actually requires that φ_1 and φ_2 are two literals built on the same predicate, say P. If $\varphi_1 = \exists P(c_1)$ and $\varphi_2 = \exists P(c_2)$, then $\exists P(c_1) \xrightarrow{\text{DEDO}} \exists P(c_2)$ simply means that $\exists P(c_2)$ can be deduced from $\exists P(c_1)$ due to specialization/generalization links (namely here, the $c_1 \xrightarrow{\text{ISA}} c_2$ link in the ontological model that relate the classes c_1 and c_2 mentioned in φ_1 and φ_2).

Technically, the $\xrightarrow{\text{DEDO}}$ links between literals are generated through a single principle:

$$
\begin{array}{lc}
\text{If in the ontology is the link} & class_1 \xrightarrow{\text{ISA}} class_2, \\[4pt]
\text{then, in the explicative model is the link} & \exists P(class_1) \xrightarrow{\text{DEDO}} \exists P(class_2) \\
\text{as well as the link} & \forall P(class_2) \xrightarrow{\text{DEDO}} \forall P(class_1) \\[4pt]
\text{Also, the following links are added} & \forall P(class_i) \xrightarrow{\text{DEDO}} \exists P(class_i).
\end{array}
\tag{2}
$$

The same principle holds for P replaced by $\neg P$. That is, from $class_1 \xrightarrow{\text{ISA}} class_2$, both $\exists \neg P(class_1) \xrightarrow{\text{DEDO}} \exists \neg P(class_2)$ and $\forall \neg P(class_2) \xrightarrow{\text{DEDO}} \forall \neg P(class_1)$ ensue, and the links $\forall \neg P(class_i) \xrightarrow{\text{DEDO}} \exists \neg P(class_i)$ are added whenever necessary.

---

[1] `FPA` stands for some precise flood-prone area, `BFPA` for the buildings in this area, `HouseFPA` for the houses in this area and `House1FPA` for the one floor low houses in this area.

Let us provide an example from Xynthia, with a predicate `Occ` so that $\exists$ `Occ(Hurri)` intuitively means: some hurricane occurs.

By means of the $\xrightarrow{\text{ISA}}$ link $\qquad$ `Hurri` $\xrightarrow{\text{ISA}}$ `SWind`,
we obtain the $\xrightarrow{\text{DEDO}}$ link $\quad \exists$ `Occ(Hurri)` $\xrightarrow{\text{DEDO}}$ $\exists$ `Occ(SWind)`.

The $\xrightarrow{\text{DEDO}}$ links between literals introduced in (2) are extended to a relation among sets of literals (as announced in Notation 3), which is done as follows:

$$
\begin{aligned}
&\text{Let } \Phi \text{ and } \Psi \quad \text{be two sets of literals,} \\
\text{we add to the explicative model} \quad & \Phi \xrightarrow{\text{DEDO}} \Psi \\
\text{if for each } \psi \in \Psi, \quad & \text{there exists } \varphi \in \Phi \text{ such that } \varphi \xrightarrow{\text{DEDO}} \psi \\
\text{and for each } \varphi \in \Phi, \quad & \text{there exists } \psi \in \Psi \text{ such that } \varphi \xrightarrow{\text{DEDO}} \psi.
\end{aligned} \tag{3}
$$

From (1), we obtain that $\psi \xrightarrow{\text{DEDO}} \psi$ is in the explicative model for each literal ψ. Accordingly,

$$
\Psi \xrightarrow{\text{DEDO}} \Psi \text{ is in the explicative model for each set of literals } \Psi. \tag{4}
$$

Back to the hurrican illustration (`Hurri` $\xrightarrow{\text{ISA}}$ `SWind`), the explicative model contains:

$\{\exists$ `Occ(Hurri)`,`ItRains`$\} \xrightarrow{\text{DEDO}} \{\exists$ `Occ(SWind)`,`ItRains`$\}$

but it does not contain

$\{\exists$ `Occ(Hurri)`,`ItRains`$\} \xrightarrow{\text{DEDO}} \{\exists$ `Occ(SWind)`$\}$

because, in the latter case, `ItRains` contributes nothing in the consequent.

Definition 4 Items (2)-(3) give all and only the ontological deduction links $\Phi \xrightarrow{\text{DEDO}} \Psi$ (introduced in Notation 3) comprised in the explicative model.

Summing up, the explicative model consists of the causal links (in the causal model) together with the ontological deduction links (obtained from the ontological model). The explicative model contains all the ingredients needed to derive explanations as is described in section 4.

2.5 Background knowledge

In addition to the explicative model, background knowledge is used for consistency issues when defining explanation links, as will be seen in Section 4, Notation 5. Background knowledge consists of logical formulas of sorted logic. Part of this knowledge is freely provided by the user. Moreover, we take causal and ontological deduction links to entail classical implication:

$$
\begin{aligned}
\{\alpha_1, \cdots, \alpha_n\} \ causes \ \beta \quad & \text{entails} \quad (\textstyle\bigwedge_{i=1}^{n} \alpha_i) \to \beta. \\
\alpha \xrightarrow{\text{DEDO}} \gamma \quad & \text{entails} \quad \alpha \to \gamma.
\end{aligned} \tag{5}
$$

Consequently, the rightmost logical formulas $(\bigwedge_{i=1}^{n} \alpha_i) \to \beta$ and $\alpha \to \gamma$ from (5), are necessarily included in the background knowledge.

3 The Xynthia Example

In this section, we consider as an example a severe storm, called Xynthia, which made 26 deaths in a single group of houses in La Faute sur Mer, a village in Vendée during a night in February 2010. It was a severe storm, with strong winds, low pressure, but it had been forecast. Since the casualties were excessive with respect to the strength of the meteorological phenomenon, various investigations have been ordered. This showed that various factors combined their effects. The weather had its role, however, other factors had been involved: recent houses and a fire station had been constructed in an area known as being susceptible of getting submerged. Also, the state authorities did not realize that asking people to stay at home was inappropriate in case of flooding given the traditionally low Vendée houses. From various enquiries, including one from the French parliament[2] and one from the Cour des Comptes (a national juridiction responsible for monitoring public accounts in France)[3] and many articles on the subject, we have plenty of information about the phenomenon and its dramatic consequences. We have extracted a small part from all this information as an illustration of our approach.

3.1 Classes and predicates for the Xynthia example

The classes we consider in the causal model are the following ones: `Hurri`, `SWind`, `BFPA`, `House1FPA`, `HouseFPA`, and `BFPA` have already been introduced in §2.3, together with a few $\xrightarrow{\text{ISA}}$ links. Among the buildings in the flood-prone area `FPA`, there is also a fire station `FireSt`. Besides `Hurri`, we consider two other kinds of natural disasters `NatDis`: tsunami `Tsun` and flooding `Flooding`. As far as meteorological phenomena are concerned, we restrict ourselves to very low pressure `VLP`, together with the aforementioned `Hurri` and `SWind`, and add high spring tide `HST` to our list of classes.

Two kinds of alerts `Alert` may be given by the authorities, Alert-Evacuate `AlertE` and Alert-StayAtHome `AlertS`. Also, `PeopleS` expresses that people stay at home. There exists an anemometer (able to measure wind strength) with a red light, described by `OK_Anemo` meaning that it is in a normal state and `Red_Anemo` meaning that its light is on, which is caused by strong wind, while during a hurricane the anemometer is in abnormal state.

The following predicates are introduced: `Flooded` and `Vict_I`, applied to a group of building, respectively meaning that flooding occurs over this group, and that there were victims in this group ($I \in \{1, 2, 3\}$ is a degree of gravity, e.g. `Vict_1`, `Vict_2` and `Vict_3` respectively mean, in % of the population of the group: at least a small number, at least a significant number and at least a large number of victims).

Remember that `Occ` means that some fact has occurred (a strong wind, a disaster, ...), similarly a unary predicate `Exp` means that some fact is expected to occur.

---

[2] http://www.assemblee-nationale.fr/13/rap-info/i2697.asp

[3] www.ccomptes.fr/Publications/Publications/Les-enseignements-des-inondations-de-2010-sur-le-littoral-atlantique-Xynthia-et-dans-le-Var

$$\exists\,\texttt{Exp(VLP)}\ \textit{causes}\ \exists\,\texttt{Exp(SWind)},$$
$$\exists\,\texttt{Occ(Hurri)}\ \textit{causes}\ \neg\,\texttt{OK\_Anemo},$$
$$\{\exists\,\texttt{Occ(SWind)},\texttt{OK\_Anemo}\}\ \textit{causes}\ \texttt{Red\_Anemo},$$
$$\{\neg\exists\,\texttt{Occ(SWind)},\texttt{OK\_Anemo}\}\ \textit{causes}\ \neg\,\texttt{Red\_Anemo},$$
$$\exists\,\texttt{Occ(NatDis)}\ \textit{causes}\ \exists\,\texttt{Occ(Alert)},$$
$$\exists\,\texttt{Exp(NatDis)}\ \textit{causes}\ \exists\,\texttt{Occ(Alert)},$$
$$\left\{\begin{array}{l}\exists\,\texttt{Occ(VLP)},\\ \exists\,\texttt{Occ(SWind)},\\ \exists\,\texttt{Occ(HST)}\end{array}\right\}\ \textit{causes}\ \exists\,\texttt{Occ(Flooding)},$$
$$\exists\,\texttt{Occ(Flooding)}\ \textit{causes}\ \forall\,\texttt{Flooded(BFPA)},$$
$$\forall\,\texttt{Flooded(BFPA)}\ \textit{causes}\ \forall\,\texttt{Vict\_1(BFPA)},$$
$$\exists\,\texttt{Occ(AlertS)}\ \textit{causes}\ \exists\,\texttt{Occ(PeopleS)},$$
$$\left\{\begin{array}{l}\exists\,\texttt{Occ(PeopleS)},\\ \forall\,\texttt{Flooded(House1FPA)}\end{array}\right\}\ \textit{causes}\ \forall\,\texttt{Vict\_2(House1FPA)},$$
$$\left\{\begin{array}{l}\forall\,\texttt{Vict\_2(House1FPA)},\\ \forall\,\texttt{Flooded(FireSt)}\end{array}\right\}\ \textit{causes}\ \forall\,\texttt{Vict\_3(House1FPA)}.$$

Table 1: Part of the causal model for Xynthia

3.2 The causal and ontological models for the Xynthia example

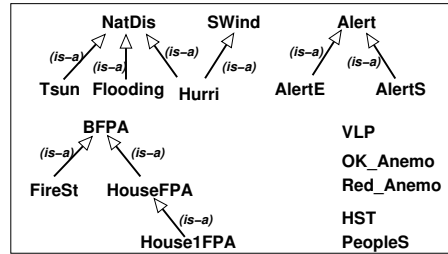The classes and the ontological model are given in Fig. 1.

Figure 1: Ontological model for Xynthia ($\overset{\text{ISA}}{\longrightarrow}$ links and constants)

Part of the causal model is given in Table 1 and in Fig. 2. It represents causal relations between (sets of) literals. It expresses that an alert occurs when a natural disaster is expected, or when a natural disaster occurs. Also, people stay at home if alerted to stay at home, and then having one level home flooded results in many victims, and even more victims if the fire station itself is flooded,...

3.3 The explicative model for the Xynthia example

The explicative model can be built, which contains various $\overset{\text{DEDO}}{\longrightarrow}$ links between literals. For instance, from $\texttt{Hurri} \overset{\text{ISA}}{\longrightarrow} \texttt{SWind}$, the links
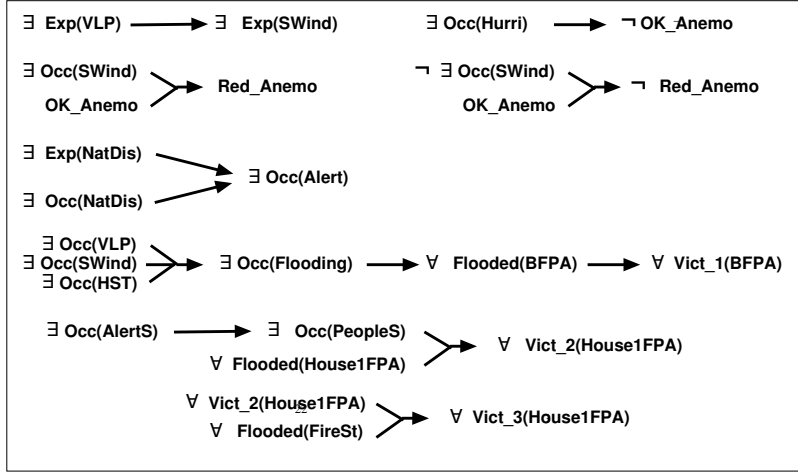
Figure 2: Part of the causal model for Xynthia

$\exists\, \text{Occ(Hurri)} \xrightarrow{\text{DEDO}} \exists\, \text{Occ(SWind)}$, and
$\exists\, \text{Exp(Hurri)} \xrightarrow{\text{DEDO}} \exists\, \text{Exp(SWind)}$ are obtained.

And similarly, from $\{\text{HouseFPA} \xrightarrow{\text{ISA}} \text{BFPA}, \ \text{House1FPA} \xrightarrow{\text{ISA}} \text{HouseFPA}\}$, we obtain $\text{House1FPA} \xrightarrow{\text{ISA}} \text{BFPA}$ by (1), from which we consequently get the link $\forall\, \text{Flooded(BFPA)} \xrightarrow{\text{DEDO}} \forall\, \text{Flooded(House1FPA)}$.

Fig. 3 represents some causal and $\xrightarrow{\text{DEDO}}$ links which are part of the explicative model. In the figures, white-headed arrows represent $\xrightarrow{\text{DEDO}}$ links. Remember that each black-headed arrow represents a causal link, from a literal or, in the case of a forked entry, from a set of literals.
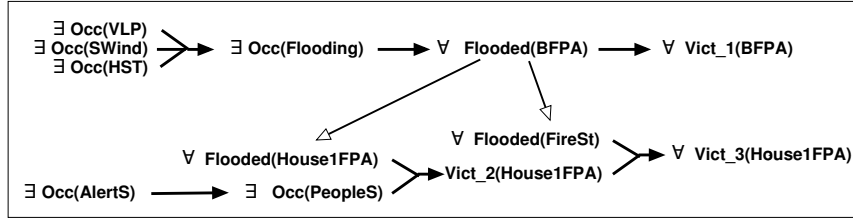

Figure 3: Part of the explicative model: data used to explain why there were numerous victims in low houses in the flood-prone area House1FPA

4 Explanations

4.1 Introducing explanation links

The explicative model (it consists of causal and ontological links) allows us to infer *explanation links*. We want to exhibit candidate reasons that can explain a fact by means of at least one causal link. We disregard "explanations" that would involve only

links which are either classical implications ($\rightarrow$) or $\xrightarrow{\text{DEDO}}$ links: some causal information is necessary for an "explanation" to hold. Here is how causal and ontological links are used in order to obtain (tentative) explanations in our formalism.

Notation 5 Let Φ, Δ and Ψ be sets of literals. An *explanation link*

$$\Phi \text{ explains } \Delta \text{ unless } \neg \Psi$$

is intended to mean that Φ explains Δ provided that, given Φ, the set Ψ is possible: if adding $\Phi \cup \Psi$ to available data (i.e., background knowledge and formulas from (5)) leads to an inconsistency, then the explanation link cannot be used to explain Δ by Φ.

Ψ is called the *provision set* of the explanation link.

When the set Ψ is empty, we may omit the "*unless* $\neg \emptyset$" (i.e., "unless $\bot$") part.

Throughout the text, we write as usual $\bigwedge \Phi$ for $\bigwedge_{\varphi \in \Phi} \varphi$ and $\neg \Psi$ for $\neg \bigwedge \Psi$.

We set the following equivalences between explanation links, so that the leftmost link can under any circumstance be substituted for the rightmost link and vice-versa:

$\Phi \text{ explains } \Delta$ is equivalent to $\Phi \text{ explains } \Delta \text{ unless } \neg \Phi$.
$\Phi \text{ explains } \Delta \text{ unless } \neg \Psi$ is equivalent to $\Phi \text{ explains } \Delta \text{ unless } \neg (\Phi \cup \Psi)$.
(6)

Let us now describe how explanation links are inferred from the explicative model. First is the case that Δ is a singleton set.

4.2 Explaining a singleton from a set of literals

The basic case consists in taking it that a direct causal link $\Phi \text{ causes } \beta$ between a set of literals Φ and a literal β provides an explanation such that the cause explains the (singleton set of) effect: see (7a).

If $\beta = \exists P(c)$ or $\beta = \neg \exists P(c)$, a more interesting case arises. Take $\beta = \exists P(c)$ for instance. Since the causal link expresses that the effect of Φ is $\exists P(c)$, it means that for any subclass c' of c, $\exists P(c')$ could be caused by Φ (unless a logical inconsistency would indicate that $\exists P(c')$ cannot be the case in the presence of Φ and background knowledge and all the formulas from (5)). Accordingly, Φ can be viewed as explaining $\exists P(c')$. This is the reason for (7b).

$$\{\Phi \text{ causes } \beta\} \quad \text{yields:} \quad \Phi \text{ explains } \{\beta\}. \quad (a)$$

$$\begin{Bmatrix} \Phi \text{ causes } \exists \beta \\ \exists \delta \xrightarrow{\text{DEDO}} \exists \beta \end{Bmatrix} \quad \text{yields:} \quad \Phi \text{ explains } \{\exists \delta\}, \text{ unless } \neg \{\exists \delta\}. \quad (b) \qquad (7)$$

(5) yields $\bigwedge \Phi \rightarrow \beta$ (case (7a)) as well as $\bigwedge \Phi \rightarrow \exists \beta$ and $\exists \delta \rightarrow \exists \beta$ (case (7b)), hence adding β (case (7a)) or $\exists \beta$ (case (7b)) to the provision set makes no difference, thereby justifying the equivalences in (6).

If $\Phi = \{\varphi\}$ is a singleton set, we may abbreviate $\Phi \text{ explains } \{\beta\}$ as $\varphi \text{ explains } \beta$.

Here are a couple of examples from the Xynthia case. First, that flooding occurred can be explained by the conjunction of very low pressure, strong wind, as well as high spring tide. In symbols,

$$\left\{ \begin{array}{l} \exists\, \texttt{Occ(VLP)}, \\ \exists\, \texttt{Occ(SWind)}, \\ \exists\, \texttt{Occ(HST)} \end{array} \right\} \; causes \; \exists\, \texttt{Occ(Flooding)}$$

yields

$$\left\{ \begin{array}{l} \exists\, \texttt{Occ(VLP)}, \\ \exists\, \texttt{Occ(SWind)}, \\ \exists\, \texttt{Occ(HST)} \end{array} \right\} \; explains \; \exists\, \texttt{Occ(Flooding)}$$

Second, expecting a hurricane can be explained from expecting very low pressure:

$$\left\{ \begin{array}{l} \exists\, \texttt{Exp(VLP)} \; causes \; \exists\, \texttt{Exp(SWind)} \\ \exists\, \texttt{Exp(Hurri)} \stackrel{\text{DEDO}}{\longrightarrow} \exists\, \texttt{Exp(SWind)} \end{array} \right\}$$

yields

$$\exists\, \texttt{Exp(VLP)} \;\; explains \; \exists\, \texttt{Exp(Hurri)}$$

Third, that all buildings in the flood-prone area are flooded can be explained by flooding:

$$\exists\, \texttt{Occ(Flooding)} \; causes \; \forall\, \texttt{Flooded(BFPA)}$$

yields

$$\exists\, \texttt{Occ(Flooding)} \;\; explains \; \forall\, \texttt{Flooded(BFPA)}$$

In the figures, dotted arrows represent explanation links (to be read *explains*), these arrows being sometimes labelled with the corresponding provision set.
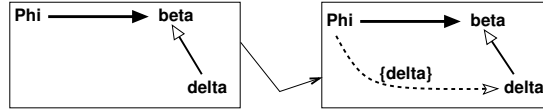

Figure 4: The schema of the explanation link from (7)

We now introduce explanation links between sets of literals, extending the notion of explanation links from sets of literals to literals presented so far. Since it is an extension, we keep the same name *explanation link*.

4.3   Explaining a set of literals from a set of literals

The patterns (7) inducing an explanation for a single observation (a singleton set) are now extended so that they can be used to obtain an explanation for a set of observations:

Let $\Phi_1, \Phi_2, \Delta, \Psi_1$ and Ψ_2 be sets of literals and β be a literal.

$$\begin{aligned} \text{If we have} \quad & \Phi_1 \; explains \; \Delta \; unless \; \neg\, \Psi_1, \quad \text{and} \\ & \Phi_2 \; explains \; \{\delta\} \; unless \; \neg\, \Psi_2, \quad\quad\quad (8) \\ \text{then we get} \quad & \Phi_1 \cup \Phi_2 \; explains \; \Delta \cup \{\delta\} \; unless \; \neg\, (\Psi_1 \cup \Psi_2). \end{aligned}$$

Notice that the condition in (8) is that $\Psi_1 \cup \Psi_2$ must be possible (it is not enough that Ψ_1 be possible **and** that Ψ_2 be possible —the same applies to (10) below).

Still further explanation links can be generated from these, *by following* $\stackrel{\text{DEDO}}{\longrightarrow}$ *links*:

$$
\text{If we have} \quad
\left\{
\begin{array}{c}
\Phi \text{ } \textit{explains } \Delta \text{ } \textit{unless } \neg \text{ } \Psi \\
\Phi_0 \stackrel{\text{DEDO}}{\longrightarrow} \Phi \\
\Delta \stackrel{\text{DEDO}}{\longrightarrow} \Delta_1
\end{array}
\right\}
\tag{9}
$$

then we get $\quad \Phi_0$ *explains* Δ_1 *unless* $\neg \text{ } \Psi$.

Let us return to our example. Applying (7a), that all the buildings in the flood-prone area are flooded can be explained by flooding (this is shown at the end of section 4.2). This explanation link (Φ is $\{\exists\,\texttt{Occ(Flooding)}\}$ and Δ is $\{\forall\,\texttt{Flooded(BFPA)}\}$) can be exploited through (9), letting $\Phi_0 = \Phi$ and $\Delta_1 = \{\forall\,\texttt{Flooded(HouseFPA)}\}$. I.e., that all houses in the flood-prone area are flooded can also be explained by flooding:

$$
\left\{
\begin{array}{c}
\exists\,\texttt{Occ(Flooding)} \textit{ causes } \forall\,\texttt{Flooded(BFPA)} \\
\forall\,\texttt{Flooded(BFPA)} \stackrel{\text{DEDO}}{\longrightarrow} \forall\,\texttt{Flooded(HouseFPA)}
\end{array}
\right\}
$$

yields

$$
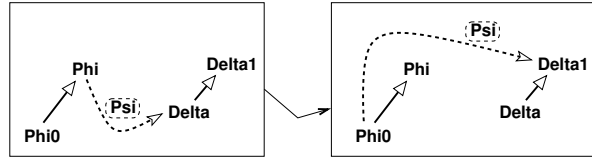\exists\,\texttt{Occ(Flooding)} \textit{ explains } \forall\,\texttt{Flooded(HouseFPA)}
$$

Figure 5: Explanation links follow $\stackrel{\text{DEDO}}{\longrightarrow}$ links [cf (9)]

The last, but not least, way by which explanation links induce further explanation links is *transitivity* (of a weak kind because provision sets are unioned).

$$
\text{If} \quad
\left\{
\begin{array}{c}
\Phi \textit{ explains } \Delta \textit{ unless } \neg \text{ } \Psi_1 \\
\Gamma \cup \Delta \textit{ explains } \Theta \textit{ unless } \neg \text{ } \Psi_2
\end{array}
\right\}
\tag{10}
$$

then $\quad \Phi \cup \Gamma$ *explains* Θ *unless* $\neg \text{ } (\Psi_1 \cup \Psi_2)$.
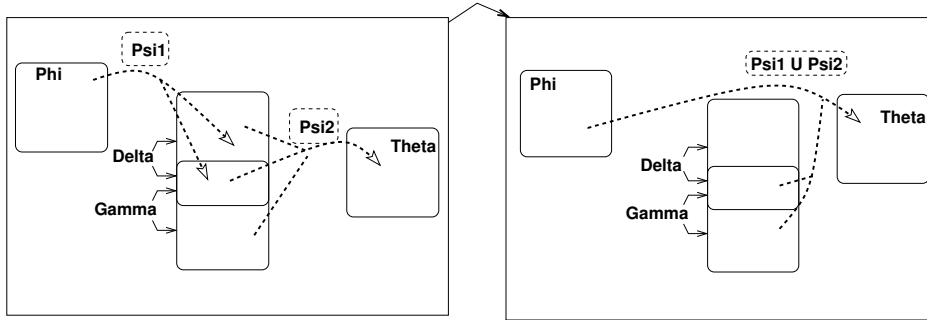
Figure 6: Transitivity of explanations among sets of literals (10)

Now, we have defined the notion introduced in Notation 5:

Definition 6 The explanation links Φ *explains* Δ *unless* $\neg \Psi$ introduced in Notation 5 arising from the explicative model are those and only those resulting from applications of (7), (8), (9) and (10).

The reader should keep in mind that Φ must always be included in the set to be checked for consistency, as is mentioned in Notation 5 (cf (6)).

Definition 6 is such that we can neither explain Φ by Φ itself nor explain Φ by Φ_0 if all we know is $\Phi_0 \stackrel{\text{DEDO}}{\longrightarrow} \Phi$. Intuitively, providing such "explanations" would be cheating, given the nature of an explanation: some causal information is required.

4.4 More examples detailed

Let us start with an example from Xynthia illustrating the use of the patterns (7b) and (9) depicted in Fig. 4 and 5.

In the causal model for Xynthia, we focus on the causal link

$$\exists\, \texttt{Exp(VLP)}\ causes\ \exists\, \texttt{Exp(SWind)}$$

In the ontological model for Xynthia, we consider the following ontological links

$$\left\{ \begin{array}{ccc} \texttt{Hurri} & \stackrel{\text{ISA}}{\longrightarrow} & \texttt{SWind} \\ \texttt{Hurri} & \stackrel{\text{ISA}}{\longrightarrow} & \texttt{NatDis} \end{array} \right\}$$

which give rise, in the explicative model, to the $\stackrel{\text{DEDO}}{\longrightarrow}$ links below

$$\left\{ \begin{array}{ccc} \exists\, \texttt{Exp(Hurri)} & \stackrel{\text{DEDO}}{\longrightarrow} & \exists\, \texttt{Exp(SWind)} \\ \exists\, \texttt{Exp(Hurri)} & \stackrel{\text{DEDO}}{\longrightarrow} & \exists\, \texttt{Exp(NatDis)} \end{array} \right\}$$

We are looking for $\texttt{Exp(NatDis)}$ to be explained by $\texttt{Exp(VLP)}$ hence we consider

$$\left\{ \begin{array}{ccc} \exists\, \texttt{Exp(VLP)} & causes & \exists\, \texttt{Exp(SWind)} \\ \exists\, \texttt{Exp(Hurri)} & \stackrel{\text{DEDO}}{\longrightarrow} & \exists\, \texttt{Exp(SWind)} \end{array} \right\}$$

and we apply (7b) in order to obtain, as a first step,

$$\exists\,\texttt{Exp(VLP)}\;\textit{explains}\;\exists\,\texttt{Exp(Hurri)}\;\textit{unless}\;\neg\,\exists\,\textit{Exp(Hurri)}$$

over which we apply (9) using the ontological deduction link obtained above, that is,

$$\exists\,\texttt{Exp(Hurri)}\;\overset{\text{DEDO}}{\longrightarrow}\;\exists\,\texttt{Exp(NatDis)}$$

in order to arrive at

$$\exists\,\texttt{Exp(VLP)}\;\textit{explains}\;\exists\,\texttt{Exp(NatDis)}\;\textit{unless}\;\neg\,\exists\,\textit{Exp(Hurri)}$$

Please observe that applying (9) actually requires $\exists\,\texttt{Exp(VLP)}\;\overset{\text{DEDO}}{\longrightarrow}\;\exists\,\texttt{Exp(VLP)}$ which is obtained by using (4).

That a natural disaster occurs can be explained from the fact that very low pressure is expected. However, if $\neg\,\exists\,\texttt{Exp(Hurri)}$ holds (it is impossible that some hurricane be expected), then this explanation no longer stands (because the effect of the causal link underlying it is strong wind and the explanation chain here identifies hurricane as the kind of strong wind expected).

Let us now turn to an example showing how a chain of explanations can be constructed by means of transitivity (10) applied over explanations already detailed above. The fact that

$$\left\{\begin{array}{l}\exists\,\texttt{Occ(VLP)},\\ \exists\,\texttt{Occ(SWind)},\\ \exists\,\texttt{Occ(HST)}\end{array}\right\}\;\textit{causes}\;\exists\,\texttt{Occ(Flooding)}$$

is in the explicative model allowed us to conclude

$$\left\{\begin{array}{l}\exists\,\texttt{Occ(VLP)},\\ \exists\,\texttt{Occ(SWind)},\\ \exists\,\texttt{Occ(HST)}\end{array}\right\}\;\textit{explains}\;\exists\,\texttt{Occ(Flooding)}\qquad\qquad\text{(i)}$$

and the fact that

$$\left\{\begin{array}{lll}\exists\,\texttt{Occ(Flooding)} & \textit{causes} & \forall\,\texttt{Flooded(BFPA)}\\ \forall\,\texttt{Flooded(BFPA)} & \overset{\text{DEDO}}{\longrightarrow} & \forall\,\texttt{Flooded(HouseFPA)}\end{array}\right\}$$

is in the explicative model allowed us to conclude

$$\exists\,\texttt{Occ(Flooding)}\;\textit{explains}\;\forall\,\texttt{Flooded(HouseFPA)}.\qquad\qquad\text{(ii)}$$

Hence chaining the explanations (i) and (ii) through (10) by letting $\Gamma=\emptyset$ yields

$$\left\{\begin{array}{l}\exists\,\texttt{Occ(VLP)}\\ \exists\,\texttt{Occ(SWind)},\\ \exists\,\texttt{Occ(HST)}\end{array}\right\}\;\textit{explains}\;\forall\,\texttt{Flooded(HouseFPA)}\qquad\qquad\text{(iii)}$$

Let us now suppose that we have multiple observations

$$\{\forall\,\texttt{Flooded(BFPA)},\texttt{Red\_Anemo}\}.$$

$$\text{From} \quad \{\exists\,\texttt{Occ(SWind)}, \texttt{OK\_Anemo}\} \quad \textit{causes} \quad \texttt{Red\_Anemo},$$
$$\text{we get} \quad \{\exists\,\texttt{Occ(SWind)}, \texttt{OK\_Anemo}\} \quad \textit{explains} \quad \texttt{Red\_Anemo}.$$

Then, from (iii), using (8) we get

$$\left\{ \begin{array}{l} \exists\,\texttt{Occ(VLP)} \\ \exists\,\texttt{Occ(SWind)}, \\ \exists\,\texttt{Occ(HST)}, \\ \texttt{OK\_Anemo} \end{array} \right\} \textit{ explains } \{\forall\,\texttt{Flooded(HouseFPA)}, \texttt{Red\_Anemo}\} \qquad \text{(iv)}$$

Also, from $\texttt{Hurri} \xrightarrow{\text{ISA}} \texttt{SWind}$ we get $\exists\,\texttt{Occ(Hurri)} \xrightarrow{\text{DEDO}} \exists\,\texttt{Occ(SWind)}$

Thus from (iii), using (9), we get

$$\left\{ \begin{array}{l} \exists\,\texttt{Occ(VLP)} \\ \exists\,\texttt{Occ(Hurri)}, \\ \exists\,\texttt{Occ(HST)} \end{array} \right\} \textit{ explains } \forall\,\texttt{Flooded(HouseFPA)} \qquad \text{(v)}$$

However, since $\quad \exists\,\texttt{Occ(Hurry)}$ *causes* $\neg\,\texttt{Red\_Anemo},\quad$ we do not get

$$\left\{ \begin{array}{l} \exists\,\texttt{Occ(VLP)} \\ \exists\,\texttt{Occ(Hurri)}, \\ \exists\,\texttt{Occ(HST)}, \\ \texttt{OK\_Anemo} \end{array} \right\} \textit{ explains } \{\forall\,\texttt{Flooded(HouseFPA)}, \texttt{Red\_Anemo}\}.$$

Fig. 7 displays another example from Xynthia of various possible explanations (represented by dotted lines) labelled as $1, 1a, \ldots$ The sets of literals, from which the explanation links start, are framed and numbered (1) to (4). These sets are not disjoint, some literals are then duplicated for readability and the copies are annotated with *(bis)*. Transitivity of explanations is again at work, e.g.,

- set 1 *explains* $\forall\,\texttt{Vict\_1(BFPA)}$ $\hspace{2cm}$ (label $1+1a+1b$)
  It is obtained by transitivity over explanation links 1, $1a$ and $1b$.

- set 4 *explains* $\forall\,\texttt{Vict\_2(House1FPA)}$ $\hspace{1.5cm}$ (label $1+1a+2$)
  It follows from explanations 1, $1a$ and 2. The latter results from explanation $1+1a$ together with the $\forall\,\texttt{Flooded(BFPA)} \xrightarrow{\text{DEDO}} \forall\,\texttt{Flooded(House1FPA)}$ link.

- set 4 *explains* $\forall\,\texttt{Vict\_3(House1FPA)}$ $\hspace{1.5cm}$ (label $1+1a+2+3$)
  Explanation 3 results from the $\forall\,\texttt{Flooded(BFPA)} \xrightarrow{\text{DEDO}} \forall\,\texttt{Flooded(FireSt)}$ link together with explanations $1+1a+2$.
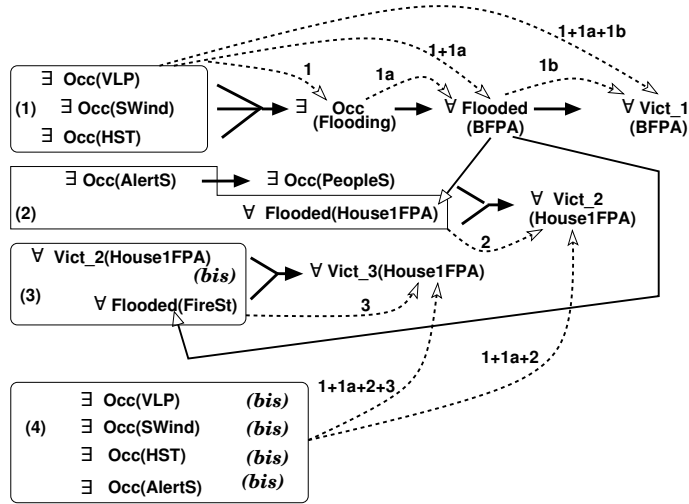
Figure 7: A few explanations for victims

5 Argumentation

The explicative causal model allows us to infer explanations for a set of statements and these explanations might be used in an argumentative context [3, 4]. Let us first provide some motivation from our case study, Xynthia.

An explanation for the flooded buildings is the conjunction of the bad weather conditions (very low pressure and strong wind) and high spring tide (see Fig. 2). Let us take this explanation as an argument. It can be attacked by noticing: a strong wind is supposed to trigger the red alarm of the anemometer and no alarm was shown. However, this counter-argument may itself be attacked by remarking that, in the case of a hurricane, that is a kind of strong wind, the anemometer is no longer operating, which explains that a red alarm cannot be observed.

Let us see how to consider formally argumentation when relying on an explicative model and explanations as described in sections 2 and 4. Of course, we begin with introducing arguments.

5.1 Arguments

An argument is a tuple $(\Phi, \Delta, \Psi, \Theta)$ such that Θ yields that

$$\Phi \text{ explains } \Delta, \text{ unless } \neg\Psi$$

is an explanation link according to Definition 6. The components of the argument are:

- Φ, *the explanation*, a set of literals.

- Δ, *the statements being explained*, a set of literals.

15

- Ψ, *the provision* of the explanation (see Section 4), a set of formulas.

- Θ, *the evidence*, comprised of formulas (e.g., $\bigwedge \Phi \to \gamma$), causal links (e.g., Φ *causes* β), and ontological deduction links (e.g., $\Delta \xrightarrow{\text{DEDO}} \{\beta\}$).

Back to (iii) in the example from Xynthia in section 4.4, that the FPA houses are flooded is explained by the set of literals

$$\Phi = \left\{ \begin{array}{c} \exists\,\texttt{Occ(VLP)} \\ \exists\,\texttt{Occ(SWind)} \\ \exists\,\texttt{Occ(HST)} \end{array} \right\}$$

on the grounds of the following set consisting of two causal links and one ontological deduction link

$$\Theta = \left\{ \begin{array}{l} \exists\,\texttt{Occ(Flooding)} \quad causes \quad \forall\,\texttt{Flooded(BFPA)}, \\ \left\{ \begin{array}{l} \exists\,\texttt{Occ(VLP)} \\ \exists\,\texttt{Occ(SWind)} \\ \exists\,\texttt{Occ(HST)} \end{array} \right\} \quad causes \quad \exists\,\texttt{Occ(Flooding)}, \\ \forall\,\texttt{Flooded(BFPA)} \quad \xrightarrow{\text{DEDO}} \quad \forall\,\texttt{Flooded(HouseFPA)} \end{array} \right\}$$

That is, (iii) gives rise to the argument $(\Phi, \{\delta\}, \Psi, \Theta)$ where

- *The explanation* is Φ.

- *There is a single statement being explained*, i.e., $\delta = \forall\,\texttt{Flooded(HouseFPA)}$.

- *The provision* of the explanation is empty.

- *The evidence* is Θ.

As for the argumentation part, our approach is concerned with sense-making. I.e., there is complex information that needs to be made sense of, and our approach is meant to provide a way to organize that information so that the key points are identified. This is a primary task in argumentation, as argumentation (even in computational guise) is much more than evaluating arguments, and in any case, does not begin with evaluating arguments [4]. Accordingly, our approach does *not* aim at evaluating a collection of arguments and counterarguments (as in the sense of determining extensions or identifying warranted arguments).

5.2 Counter-arguments

A counter-argument for an argument $(\Phi, \Delta, \Psi, \Theta)$ is an argument $(\Phi', \Delta', \Psi', \Theta')$ which questions

1. either Φ (e.g., an argument exhibiting an explanation for $\neg\Phi$)

2. or Δ (e.g., an argument exhibiting an explanation for $\neg\Delta$)

3. or Ψ (e.g., an argument exhibiting an explanation for $\neg\Psi$)

4. or any item in Θ (e.g., an argument exhibiting an explanation for the negation of some θ occurring in Θ)

5. or does so by refutation: i.e an argument exhibiting an explanation for a statement known to be false and using any of Φ, Θ, Ψ and Δ. In this case, at least one of Φ', Θ', Ψ' intersects one of Φ, Δ, Θ, or Ψ.

Type (5) counter-arguments do not directly oppose an item in the argument being challenged. They rather question such an item by using it to provide an argument whose conclusion is wrong. The presence of such an item is ensured by checking that the challenged argument and the counter-argument indeed share something in common, i.e., that the intersection is not empty. Otherwise, in the case that the intersection is empty, then the two arguments have nothing in common, hence none can be viewed as a counter-argument to the other.

These counter-arguments have the form of an argument. They explain something that contradicts something in the challenged argument.

Dispute.
Let us consider the illustration at the start of this section: The argument (that the houses in the flood-prone area are flooded is partly explained by a strong wind) is under attack on the grounds that the anemometer did not turn red – indicating that no strong wind occurred. The latter is a counter-argument of type 5 in the above list. Indeed, the statement explained by the counter-argument is `Red_Anemo` that has been observed to be false. The explanation uses $\exists$ `Occ(SWind)`, i.e., an item used by the explanation and then belonging to Φ in the attacked argument.

Taking `Red_Anemo` to be a falsehood, the counter-argument $(\Phi', \Delta', \emptyset, \Theta')$ results from Θ' yielding that

$$
\left\{ \begin{array}{c} \exists\, \texttt{Occ(SWind)} \\ \texttt{OK\_Anemo} \end{array} \right\} \quad \text{explains} \quad \{\texttt{Red\_Anemo}\}
$$

where

- *The explanation* is
$$
\Phi' = \{\exists\, \texttt{Occ(SWind)}, \texttt{OK\_Anemo}\}
$$

- *The statement being explained* is
$$
\Delta' = \{\texttt{Red\_Anemo}\}
$$

- *The evidence* is
$$
\Theta' \quad = \quad \left\{ \left\{ \begin{array}{c} \exists\, \texttt{Occ(SWind)} \\ \texttt{OK\_Anemo} \end{array} \right\} \text{\textit{causes} } \texttt{Red\_Anemo} \right\}
$$

Notice that Φ' does intersect Φ.

This is a counter-argument because, taking the anemometer being red as falsity, it is an argument which uses the occurring of a strong wind to conclude the anemometer being red. As explained above in the general case, such a type (5) counter-argument uses an item (a strong wind occurring) from the argument being challenged, in order to conclude a falsity (the anemometer being red).

Dispute (continued)

This counter-argument has in turn a counter-argument (of type 1.). It explains the misbehavior of the anemometer by the occurrence of an hurricane (that is a strong wind), and then explains the negation of an item OK_Anemo of the explanation Φ' of the counter-argument. The anemoter not getting red, instead of being explained by the absence of a strong wind, is explained by the fact that the wind was so strong (an hurricane) that the anemometer misbehaved.

So, the counter-counter-argument is: $(\Phi'', \Delta'', \emptyset, \Theta'')$ resulting from Θ'' yielding that:

$$\{ \ \exists \, \texttt{Occ(Hurri)} \ \} \text{ explains } \{\neg \texttt{OK\_Anemo}\}$$

where

- *The explanation* is
$$\Phi'' = \{\exists \, \texttt{Occ(Hurri)}\}$$

- *The statement being explained* is
$$\Delta'' = \{\neg \texttt{OK\_Anemo}\}$$

- *The evidence* is
$$\Theta'' = \{\exists \, \texttt{Occ(Hurri)} \ causes \ \neg \texttt{OK\_Anemo}\}$$

The dispute can extend to a counter-counter-counter-argument and so on as the process iterates.

6 Conclusion

The contribution of our work is firstly to propose a new logic-based formalism where explanations result from both causal and ontological links. It is important to stress that our approach reasons from causal relationships which are given, in contrast to a number of models for causality that aim at finding causal relationships (e.g., [8, 9]). This causal-based approach for explanations, already defended in [1], is relatively different from other work on explanations that rely on expert knowledge and are considered as useful functionalities for expert systems and recommender systems (for a synthetic view on explanations in these domains, see [5, 10, 14]. We then show how these explanation links may be interestingly used as building blocks in an argumentative context [3]. It has similarities with the work by [12], who argue that, in the context of knowledge-based decision support systems, integrating explanations and argumentation capabilities is a valuable perspective.

Although explanation and argumentation have long been identified as distinct processes [13], it is also recognized that the distinction is a matter of context, hence they both play a role [7] when it comes to eliciting an answer to a "why" question. This is exactly what is attempted in this paper, as we are providing "possible" explanations, that thus can be turned into arguments. The argument format has some advantages inasmuch as its uniformity allows us to express objection in an iterated way: "possible" explanations are challenged by counter-arguments that happen to represent rival, or incompatible, "possible" explanations. Some interesting issues remain to be

studied. Among others, comparing competing explanations according to minimality, preferences, and generally a host of criteria.

We have designed a system in answer set programming that implements the explicative proposal introduced above. Indeed, answer set programming [6] is well fitted for this kind of problem. One obvious reason is that rules such as (5), (8) or (9) can be translated literally and efficiently. Also ASP is known to be good for working with graphs such as the one depicted in the figures of this text. We plan to include our system in an argumentative framework and think it will be a good basis for a really practical system, able to manage with as a rich and tricky example as the Xynthia example.

Acknowledgements

References

[1] Philippe Besnard, Marie-Odile Cordier, and Yves Moinard. Deriving explanations from causal information. In Malik Ghallab and Constantine D. Spytopoulos and Nikos Fakotakis and Nikolaos M. Avouris, editor, *ECAI 2008*, pages 723–724. IOS Press, 2008.

[2] Philippe Besnard, Marie-Odile Cordier, and Yves Moinard. Ontology-based inference for causal explanation. *Integrated Computer-Aided Engineering*, 15:351–367, 2008.

[3] Philippe Besnard and Anthony Hunter. *Elements of Argumentation*. MIT Press, Cambridge, 2008.

[4] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[5] Gerhard Friedrich and Markus Zanker. A taxonomy for generating explanations in recommender systems. *AI Magazine*, 32(3):90–98, 2011.

[6] Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. *Answer Set Solving in Practice*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers, 2012.

[7] Justin Scott Giboney, Susan Brown, and Jay F. Nunamaker Jr. User acceptance of knowledge-based system recommendations: Explanations, arguments, and fit. In *45th Annual Hawaii International Conference on System Sciences (HICSS'45)*, pages 3719–3727. IEEE Computer Society, 2012.

[8] Joseph Halpern and Judea Pearl. Causes and Explanations: A Structural-Model Approach. Part I: Causes. In Jack S. Breese and Daphne Koller, editors, *17th Conference in Uncertainty in Artificial Intelligence (UAI'01)*, pages 194–202. Morgan Kaufmann, 2001.

[9] Joseph Y. Halpern and Judea Pearl. Causes and Explanations: A Structural-Model Approach. Part II: Explanations. In Bernhard Nebel, editor, *17th International Joint Conference on Artificial Intelligence (IJCAI'01)*, pages 27–34. Morgan Kaufmann, 2001.

[10] Carmen Lacave and Francisco Javier Díez. A review of explanation methods for heuristic expert systems. *The Knowledge Engineering Review*, 19:133–146, 6 2004.

[11] Dov Hugh Mellor. *The Facts of Causation*. Routledge, London, 1995.

[12] Bernard Moulin, Hengameh Irandoust, Micheline Bélanger, and Gaëlle Desbordes. Explanation and argumentation capabilities: Towards the creation of more persuasive agents. *Artif. Intell. Rev.*, 17(3):169–222, May 2002.

[13] Douglas Walton. Explanations and arguments based on practical reasoning. In Thomas Roth-Berghofer, Nava Tintarev, and David B. Leake, editors, *Workshop on Explanation-Aware Computing at IJCAI'09*, pages 72–83, Pasadena, CA, U.S.A., July 2009.

[14] L. Richard Ye and Paul E. Johnson. The impact of explanation facilities in user acceptance of expert system advice. *MIS Quarterly*, 19(2):157–172, 1995.