Ontology-based inference for causal explanation

Philippe Besnard, Marie-Odile Cordier, Yves Moinard

Ontology-based inference for causal explanation *

Ph. Besnard[1] M.-O. Cordier[2 4] Y. Moinard[3 4]

May 6, 2008

Abstract

We define an inference system to capture explanations based on causal statements, using an ontology in the form of an $IS\text{-}A$ hierarchy. We first introduce a simple logical language which makes it possible to express that a fact causes another fact and that a fact explains another fact. We present a set of formal inference patterns from causal statements to explanation statements. We introduce an elementary ontology which gives greater expressiveness to the system while staying close to propositional reasoning. We provide an inference system that captures the patterns discussed, firstly in a purely propositional framework, then in a datalog (limited predicate) framework.

1 Introduction

We are aiming at a logical formalization of explanations from causal statements. For example, it is usually admitted that fire is an explanation for smoke, on the grounds that fire causes smoke. In other words, fire causes smoke is a premise from which it can be inferred that fire is an explanation for smoke. In this particular example, concluding from cause to explanation is immediate but such is not always the case, far from it. In general, the reasoning steps leading from cause to explanation are not so trivial:

Example. *We consider two causal statements:*
(i) Any ship that is about to sink causes her crew to launch some red rocket(s)
(ii) On July the 14^{th}*, the celebration of the French national day causes the launching of fireworks all over France.*
So, if the place is a coastal city in France, on July the 14^{th}*, then red rockets*

* This preprint of a paper published in Integrated Computer-Aided Engineering Journal (publisher: IOS Press, editor: Hojjat Adeli), 15(4), pp. 351–367, 2008) is an extended version of [BCM07].

[1]CNRS, IRIT, Université Paul Sabatier, 31062 Toulouse cedex, France, besnard@irit.fr
[2]Institution: Université de Rennes I
[3]Institution: INRIA
[4]Address: IRISA, Campus de Beaulieu, 35042 Rennes cedex, France, {cordier,moinard}@irisa.fr

being launched could be explained either by some ship(s) sinking or by a national day firework launched.

In this example, it is needed to acknowledge the fact that
a red rocket is a kind of (colourful) rocket
in order to get the second explanation, which makes sense.

Example (cont'd). *Suppose that we now add the following statement:*
(i) Seeing a red rocket being launched triggers a rescue process.
Now, on July the 14^{th} *in a coastal city in France, a possible explanation for the triggering of the rescue process, as happens in practice, is that a national day firework has been launched.*

Thus we say that *"*α *explains* β*"* when adding α to our knowledge, and using a "suitable chain" of causal and taxonomical information, β is obtained. Which chains are "suitable" is one of the subjects addressed in this text. We define a dedicated inference system to capture explanations based on causal statements and stress that the rôle of ontology-based information is essential. Our causal information is restricted to the cases where the causation never fails, rejecting e.g. "smoking causes cancer". We leave also for future work temporal aspects. Also, we consider that the causal information is provided by the user, we are not concerned by the extraction of causal information as in scientific research. We provide a way to extract what we call explanations from causal (and "ontological") information given by the user: we aim at providing all the (eventually tentative) explanations that can be obtained. Then, some choice between these explanations should be made by the user, depending of its needs, but this aspect is not considered here.

In the second section, we introduce the propositional logical language that we propose to use, then we define the set of patterns dedicated to inferring explanations from causal statements and ontological information. In the third section we extend the formalism to a restricted predicate case (a la "datalog", no quantifiers admitted in the formulas), the ontology consisting in links between constant symbols. We introduce two kinds of parameters for a predicate: "existential" and "universal" ones. Then we extend this to predicates of any arity and we introduce also ontological links between predicates. In the fourth section, we discuss a few features of the formalism. In the conclusion, we summarize the main points and we propose some possible future work.

2 The propositional formalism

2.1 Vocabulary and first properties

For the sake of clarity, we present the propositional version of the formalism first. We distinguish various types of statements in our formal system:

C: A theory expressing causal statements. E.g. On_alarm *causes* $Heard\_bell$ or Flu *causes* $Fever\_Temperature$.

O: An ontology in the form of a set of $IS\text{-}A$ links between two items which can appear in a causal statement.
E.g., $Temperature\_39 \rightarrow_{IS-A} Fever\_Temperature$,
$Temperature\_41 \rightarrow_{IS-A} Fever\_Temperature$,
$Heard\_loud\_bell \rightarrow_{IS-A} Heard\_bell$,
$Heard\_soft\_bell \rightarrow_{IS-A} Heard\_bell$.

W: A classical propositional theory expressing truths (i.e., incompatible facts, co-occurring facts, ...). E.g., $Heard\_soft\_bell \rightarrow \neg Heard\_loud\_bell$).

Intuitively, propositional symbols denote elementary properties describing states of affairs, which can be "facts" or "events" such as $Fever\_Temperature$, On_alarm, $Heard\_bell$.

The causal statements express causal relations between facts or events expressed by these propositional symbols. Some care is necessary when providing these causal and ontological atoms. If "Flu *causes* $Fever\_Temperature$", we will conclude Flu *explains* $Temperature\_39$ from $Temperature\_39 \rightarrow_{IS-A} Fever\_Temperature$, but we cannot state Flu *causes* $Temperature\_39$: we require that the causal information is provided "on the right level" and in this case, $Temperature\_39$ is not on the right level.

Besides, our restricted ontology could be termed "taxonomy".

The formal system we introduce below is meant to infer, from such premises $C \cup O \cup W$, formulas denoting explanations. This inference will be denoted $\vdash_C$. The ontological atoms express some common sense knowledge which is necessary to infer these "explanations". Notice that a feature of our formalism is that standard implication alone cannot help to infer explanations [BCM06, BCM07].

In this section, $\alpha, \beta, \ldots$ denote the propositional atoms and $\Phi, \Psi, \ldots$ denote sets thereof.

Atoms

1. *Propositional atoms*: $\alpha, \beta, \ldots$.
2. *Causal atoms*: α *causes* β.
3. *Ontological atoms*: $\alpha \rightarrow_{IS-A} \beta$.
4. *Explanation atoms*: α *explains* β *because_possible* Φ.

An ontological atom reads: α is a β.
An explanation atom reads: α *is an explanation for* β *because* Φ *is possible.*

Notation: In order to help reading long formulas, explanation atoms are sometimes abbreviated as α *explains* β *bec_poss* Φ.

Formulas

1. *Propositional formulas*: Boolean combinations of propositional atoms.
2. *Causal formulas*: Boolean combinations of causal or propositional atoms.

The premises of the inference $\vdash_C$, namely $C \cup O \cup W$, consist of propositional and causal formulas, and ontological *atoms* (no ontological formula). Notice that explanation atoms cannot occur in the premises.

The properties of causal and ontological formulas we consider are as follows.

1. **Properties of the causal operator**
   (a) *Entailing [standard] implication*: If α *causes* β, then $\alpha \rightarrow \beta$.
2. **Properties of the ontological operator**
   (a) *Entailing implication*: If $\alpha \rightarrow_{IS-A} \beta$, then $\alpha \rightarrow \beta$.
   (b) *Transitivity*: If $a \rightarrow_{IS-A} b$ and $b \rightarrow_{IS-A} c$, then $a \rightarrow_{IS-A} c$.
   (c) *Reflexivity*: $c \rightarrow_{IS-A} c$.

Reflexivity is an unconventional property for an *IS-A* hierarchy. It is included here because it helps keeping the number of inference schemes low (see later).

W is supposed to include (whether explicitly or via inference) all the implications induced by the ontological atoms. For example, if $Heard\_loud\_bell \rightarrow_{IS-A} Heard\_bell$ is in O then $Heard\_loud\_bell \rightarrow Heard\_bell$ is in W. Similarly, W is supposed to include all conditionals induced by the causal statements in C. For example, if Flu *causes* $Fever\_Temperature$ is in C, then $Flu \rightarrow Fever\_Temperature$ is in W.

2.2 Patterns for inferring explanations

A set of patterns, introduced in [BCM07], is proposed to infer explanations from premises $C \cup O \cup W$. Before providing the rules (see § 2.3 below), let us motivate these rules by listing the main patterns that we consider as desirable. The base case explains β from α whenever α *causes* β (§ 2.2.1). More elaborate cases explain β from α whenever α *causes* some β' ontologically related with β (§ 2.2.2, 2.2.3). Finally, explanations should be transitive (almost) (§ 2.2.4).

2.2.1 The base case

A basic idea is that what causes an effect can always be suggested as an explanation when the effect happens to be the case:

$$\left\{ \begin{array}{ll} \text{If} & \alpha \textit{ causes } \beta \\ \text{and} & W \not\models \neg\alpha \end{array} \right\} \quad \text{then} \quad \alpha \textit{ explains } \beta \textit{ because\_possible } \{\alpha\}.$$

Example. *Consider a causal model such that* $W \not\vdash \neg Flu$ *and* O *is empty whereas* $C = \{Flu \textit{ causes } Fever\_Temperature\}$.

Then, the atom Flu *explains* $Fever\_Temperature$ *because_possible* $\{Flu\}$ *is inferred. That is,* Flu *is an explanation for* $Fever\_Temperature$.

Notice that since $Flu \rightarrow Fever\_Temperature$ is in W, we get in fact that $W \not\vdash \neg Flu$ is equivalent to $W \not\vdash \neg(Flu \wedge Fever\_Temperature)$ which is why $Fever\_Temperature$ is not included in the set of the "conditions" for this explanation.

By the way, "is an explanation" must be understood as provisional. Inferring that Flu is an explanation for $Fever\_Temperature$ is a tentative conclusion: Should Flu be ruled out, e.g., $\neg Flu \in W$, then Flu is no longer an explanation for $Fever\_Temperature$.
Formally, with $Form = Flu$ *explains* $Fever\_Temperature$ *bec_poss* $\{Flu\}$,
we get $C \cup O \cup W \vdash_C Form$; but $C \cup O \cup W \cup \{\neg Flu\} \not\vdash_C Form$.

2.2.2 Wandering the IS-A hierarchy: Going upward

What causes an effect can be suggested as an explanation for any consistent ontological generalization of the effect:

$$\left\{ \begin{array}{ll} \text{If} & \alpha \textit{ causes } \beta, \\ & \beta \rightarrow_{IS-A} \gamma, \\ \text{and} & W \not\models \neg\alpha \end{array} \right\} \quad \text{then} \quad \alpha \textit{ explains } \gamma \textit{ because\_possible } \{\alpha\}.$$

Example. $C = \{On\_alarm \textit{ causes } Heard\_bell\}$
$O = \{Heard\_bell \rightarrow_{IS-A} Heard\_noise\}$

O states that hearing a bell is more precise than hearing a noise. Since On_alarm is an explanation for $Heard\_bell$ from the base case, it is also an explanation for $Heard\_noise$. Let the causal theory CT consist in the two preceding atoms of C and O, W containing nothing else than the implications induced by C and O, that is:

$$W = \left\{ \begin{array}{c} On\_alarm \rightarrow Heard\_bell, \\ Heard\_bell \rightarrow Heard\_noise \end{array} \right\}$$

We get $CT \vdash_C On\_alarm$ explains $Heard\_noise$ bec_poss$\{On_alarm\}$.

Then, we additionally know that hearing a fog-horn is more precise than hearing a noise, that a fog-horn is heard, and that hearing a fog-horn is not hearing a bell. This is expressed by the causal theory CT', defined by the sets C as above, $O \cup O'$ and $W \cup W'$ with O' and W' as follows (W' contains the new implication induced by O', plus the other additional information):

$$O' = \{Heard\_fog\text{-}horn \rightarrow_{IS-A} Heard\_noise\}.$$

$$W' = \left\{ \begin{array}{c} Heard\_fog\text{-}horn \rightarrow Heard\_noise, \\ Heard\_fog\text{-}horn, \\ \neg(Heard\_bell \leftrightarrow Heard\_fog\text{-}horn) \end{array} \right\}$$

Even taking into account the fact that $Heard\_bell$ is an instance of $Heard\_noise$, it can no longer be inferred that On_alarm is an explanation for $Heard\_noise$:

$CT' \not\vdash_C On\_alarm$ *explains* $Heard\_noise$ *bec_poss*$\{On\_alarm\}$.

The inference fails because it would need $Heard\_noise$ *to be of the* $Heard\_bell$ *kind (which is false, cf* $Heard\_fog\text{-}horn$*). Technically, the inference fails because* $W \cup W' \vdash \neg On\_alarm$.

The next example illustrates why resorting to ontological information is essential when attempting to infer explanations: the patterns in the present § 2.2.2 as well as in following § 2.2.3 extend the base case for explanations to *ontology-based* consequences, not to any consequences.

Example. *Rain makes me growl. Trivially, I growl only if I am alive. However, rain cannot be taken as an explanation for the fact that I am alive.*

$C = \{Rain\ causes\ I\_growl\}, \quad O = \emptyset, \quad W = \{I\_growl \rightarrow I\_am\_alive\}.$
We get: $C \cup O \cup W \not\vdash_C$ $Rain$ *explains* I_am_alive *because_possible* $\{Rain\}$

2.2.3 Wandering the IS-A hierarchy: Going downward

What causes an effect can presumably be suggested as an explanation when the effect takes place in one of its specialized forms:

$$\left\{ \begin{array}{ll} \text{If} & \alpha\ causes\ \beta, \\ & \gamma \rightarrow_{IS-A} \beta, \\ \text{and} & W \not\models \neg(\alpha \wedge \gamma), \end{array} \right\} \quad \text{then} \quad \alpha\ \ explains\ \ \gamma\ \ because\_possible\ \{\alpha, \gamma\}.$$

Example. *Consider a causal model with* C *and* O *as follows:*

$C = \{On\_alarm\ causes\ Heard\_bell\}$ *and*

$$O = \left\{ \begin{array}{l} Heard\_loud\_bell \rightarrow_{IS-A} Heard\_bell \\ Heard\_soft\_bell \rightarrow_{IS-A} Heard\_bell \end{array} \right\}$$

O *means that* $Heard\_loud\_bell$ *and* $Heard\_soft\_bell$ *are more precise than* $Heard\_bell$.

Since On_alarm *is an explanation for* $Heard\_bell$*, it also is an explanation for* $Heard\_loud\_bell$ *and similarly* $Heard\_soft\_bell$*. This holds inasmuch as there is no statement to the contrary:*

The latter inference would not be drawn if for instance $\neg Heard\_soft\_bell$ *or* $\neg(Heard\_soft\_bell \wedge On\_alarm)$ *were in* W.

Formally, with $(Form\_loud) =$
On_alarm *explains* $Heard\_loud\_bell$ *bec_poss* $\{On\_alarm, Heard\_loud\_bell\}$
and $(Form\_soft) =$
On_alarm *explains* $Heard\_soft\_bell$ *bec_poss* $\{On\_alarm, Heard\_soft\_bell\}$*:*

$C \cup O \cup W \ \vdash_C \ (Form\_loud), \qquad C \cup O \cup W \ \vdash_C \ Form\_soft, \qquad$ *and*
$C \cup O \cup W \cup \{\neg(Heard\_soft\_bell \wedge On\_alarm)\} \ \not\vdash_C \ (Form\_soft)$

Here there is another example

Example. $C = \{Flu\ causes\ Fever\_Temperature\}$ *and*
$O = \{Temperature\_39 \rightarrow_{IS-A} Fever\_Temperature\}$.
W *contains no statement apart from those induced by* C *and* O*, that is:*

$$W = \{Flu \rightarrow Fever\_Temperature, Temperature\_39 \rightarrow Fever\_Temperature\}$$

Inasmuch as $Fever\_Temperature$ *could be* $Temperature\_39$*,* Flu *then counts as an explanation for* $Temperature\_39$*.*

$$C \cup O \cup W \vdash_C Flu\ explains\ Temperature\_39\ bec\_poss\ \{Flu, Temperature\_39\}$$

Again, it would take $Flu \wedge Temperature\_39$ *to be ruled out for the inference to be prevented.*

2.2.4 Transitivity of explanations

We make no assumption as to whether the causal operator is transitive (from α *causes* β and β *causes* γ does α *causes* γ follow?). However, we do regard inference of explanations as transitive which, in the simplest case, means that if α explains β and β explains γ then α explains γ. Notice already that, since this transitivity of explanations "gathers the conditions" (Point 3b § 2.3 below), it is not absolute, and can easily be blocked.

The general pattern for transitivity of explanations takes two causal statements, α *causes* β_1 and β_2 *causes* γ where β_1 and β_2 are ontologically related, as premises in order to infer that α is an explanation for γ.

In the first form of transitivity, β_2 is inherited from β_1 by going upward in the IS-A hierarchy.

$$\left\{\begin{array}{ll} \text{If} & \alpha\ causes\ \beta_1,\ \ \beta_2\ causes\ \gamma, \\ & \beta_1 \rightarrow_{IS-A} \beta_2, \\ \text{and} & W \not\models \neg\alpha, \end{array}\right\} \quad \text{then} \quad \alpha\ \ explains\ \ \gamma\ \ bec\_poss\{\alpha\}.$$

Example. *Sunshine makes me happy. Being happy is why I sing. Therefore, sunshine is a plausible explanation for the case that I am singing.*

$$C = \left\{\begin{array}{c} Sunshine\ causes\ I\_am\_happy \\ I\_am\_happy\ causes\ I\_am\_singing \end{array}\right\}$$
$$W = \left\{\begin{array}{c} Sunshine \rightarrow I\_am\_happy \\ I\_am\_happy \rightarrow I\_am\_singing \end{array}\right\}$$

So, for the inference relation $\vdash_C$*,* $C \cup O \cup W$ *infers the atom:*

$$Sunshine\ \ explains\ \ I\_am\_singing\ \ because\_possible\ \{Sunshine\}.$$

The above example exhibits transitivity of explanations for the simplest case that $\beta_1 = \beta_2$ in the pattern α *causes* β_1 and β_1 *causes* γ entail
α *explains* γ *because_possible* $\{\alpha\}$ (trivially, if $\beta_1 = \beta_2$ then $\beta_1 \rightarrow_{IS-A} \beta_2$).

This is one illustration that using reflexivity in the ontology relieves us from the burden of tailoring definitions to capture formal degenerate cases.

The next example exhibits the general case $\beta_1 \neq \beta_2$ in the pattern given above.

Example. *Let* $O = \{Heard\_bell \rightarrow_{IS-A} Heard\_noise\}$ *and*

$$C = \left\{ \begin{array}{c} On\_alarm\ causes\ Heard\_bell \\ Heard\_noise\ causes\ Disturbance \end{array} \right\}$$

W *states the facts induced by* C *and* O*, that is:*

$$W = \left\{ \begin{array}{c} On\_alarm \rightarrow Heard\_bell \\ Heard\_noise \rightarrow Disturbance \\ Heard\_bell \rightarrow Heard\_noise \end{array} \right\}$$

So, for the inference relation $\vdash_C$*,* $C \cup O \cup W$ *infers the atom:*

$$On\_alarm\ explains\ Disturbance\ bec\_poss\ \{On\_alarm\}.$$

In the second form of transitivity, β_1 is inherited from β_2 by going downward in the IS-A hierarchy.

$$\left\{ \begin{array}{ll} \text{If} & \alpha\ causes\ \beta_1,\ \ \beta_2\ causes\ \gamma, \\ & \beta_2 \rightarrow_{IS-A} \beta_1, \\ \text{and} & W \not\models \neg(\alpha \wedge \beta_2), \end{array} \right\} \text{then}\quad \alpha\ \ explains\ \ \gamma\ \ bec\_poss\{\alpha, \beta_2\}.$$

Example. $O = \{Heard\_loud\_bell \rightarrow_{IS-A} Heard\_bell\}$

$$C = \left\{ \begin{array}{c} On\_alarm\ causes\ Heard\_bell \\ Heard\_loud\_bell\ causes\ Deafening \end{array} \right\}$$

$$W = \left\{ \begin{array}{c} Heard\_loud\_bell \rightarrow Heard\_bell \\ On\_alarm \rightarrow Heard\_bell \\ Heard\_loud\_bell \rightarrow Heard\_Deafening \end{array} \right\}$$

On_alarm does not ***cause*** $Heard\_loud\_bell$ *(neither does it* ***cause*** $Deafening$*), but it is an* ***explanation*** *for* $Heard\_loud\_bell$ *by virtue of the downward scheme. Due to the base case,* $Heard\_loud\_bell$ *is in turn an explanation for* $Deafening$*. In fact, On_alarm is an explanation for* $Deafening$ *by virtue of transitivity.*

Considering a causal operator which is transitive would give the same explanations but is obviously more restrictive as we may not want to endorse an account of causality which is transitive. Moreover, transitivity for explanations not only seems right in itself but it also means that our model of explanations can be plugged with any causal system whether transitive or not.

The preceding examples are here to introduce the general pattern for transitivity of explanations which is as follows:

$$\left\{ \begin{array}{ll} \text{If} & \alpha\ explains\ \beta\ bec\_poss\ \Phi, \\ & \beta\ explains\ \gamma\ bec\_poss\ \Psi, \\ \text{and} & W \not\models \neg \bigwedge(\Phi \cup \Psi), \end{array} \right\} \quad \text{then} \quad \alpha\ explains\ \gamma\ bec\_poss\ \Phi \cup \Psi.$$

2.2.5 Explanation provisos and their simplifications

Explanation atoms are written α *explains* β *because_possible* Φ as the definition is intended to make the atom true just in case it is successfully checked that the proviso is possible: An explanation atom is not to be interpreted as a kind of conditional statement. Indeed, we do not write *"if_possible"*. The argument in *"because_possible"* gathers those conditions that must be possible together if α is to explain β (there can be others: α can also be an explanation of β with respect to another set of arguments in *"because_possible"*).

Notice that the set of conditions in an explanation atom can often be simplified. Using $\bigwedge \Phi$ to denote the conjunction of the formulas in the set Φ, the following general scheme amounts to simplifying the proviso attached to an explanation atom.

$$\left\{\begin{array}{ll} \text{If} & \text{for all } i \in \{1, \cdots, n\},\ \alpha \textit{ explains } \beta \textit{ because\_possible } (\Phi_i \cup \Phi), \\ \text{and} & W \models \bigwedge \Phi \rightarrow \bigvee_{i=1}^{n} \bigwedge \Phi_i, \end{array}\right\}$$

$$\text{then} \qquad \alpha \textit{ explains } \beta \textit{ because\_possible } \Phi.$$

As a simple motivating example for this general scheme, let us consider the case where we have

$$\alpha \textit{ causes } \beta \quad \textit{and} \quad \beta \textit{ causes } \gamma.$$

Then we get α *explains* β *bec_poss*$\{\alpha\}$ and β *explains* γ *bec_poss*$\{\beta\}$ by § 2.2.1, thus α *explains* γ *bec_poss*$\{\alpha, \beta\}$ by the general pattern for transitivity § 2.2.4. Now, we get $\alpha \rightarrow \beta$ from α *causes* β, thus, $W \models \alpha$ is equivalent to $W \models \alpha \wedge \beta$: the two sets of condition $\{\alpha\}$ and $\{\alpha, \beta\}$ are equivalent here, which justifies to simplify the explanation atom into α *explains* γ *because_possible* $\{\alpha\}$. This provides a justification for the general pattern of simplification of the set of condition where $n = 1, \Phi = \{\alpha\}$ and $\Phi_1 = \{\beta\}$.

More complex examples may involve disjunctions, such as in the small following example:

Let us suppose that we have derived the following two explanation atoms: α *explains* γ *bec_poss*$\{\alpha, \beta_1\}$ and α *explains* γ *bec_poss*$\{\alpha, \beta_2\}$. Let us suppose that W contains $\alpha \rightarrow (\beta_1 \vee \beta_2)$. Then, as soon as α, together with either β_1 or β_2, is possible, we get that α explains γ. Now, $[W \not\models \neg(\alpha \wedge \beta_1)$ or $W \not\models \neg(\alpha \wedge \beta_2)]$ is equivalent to $[W \not\models \neg(\alpha \wedge (\beta_1 \vee \beta_2))]$ and, from $\alpha \rightarrow (\beta_1 \vee \beta_2)$, this is equivalent to $[W \not\models \neg\alpha]$.

Thus, it is natural to get the explanation atom α *explains* γ *bec_poss*$\{\alpha\}$. This is the general pattern for simplification of the conditions of explanation where $n = 2$, $\Phi = \{\alpha\}$, $\Phi_1 = \{\beta_1\}$ and $\Phi_2 = \{\beta_2\}$. Generalizing this example produces naturally the general pattern for simplification of the conditions.

2.3 A formal system for inferring explanations

The above ideas are embedded in a short proof system extending classical logic:

1. ***Causal formulas*** $(\alpha \; causes \; \beta) \rightarrow (\alpha \rightarrow \beta)$.

2. ***Ontological atoms***

   (a) If $\beta \rightarrow_{IS-A} \gamma$ then $\beta \rightarrow \gamma$.

   (b) If $\alpha \rightarrow_{IS-A} \beta$ and $\beta \rightarrow_{IS-A} \gamma$ then $\alpha \rightarrow_{IS-A} \gamma$.

   (c) $\alpha \rightarrow_{IS-A} \alpha$

3. ***Explanation atoms***

   (a) *Base case*
   If $\delta \rightarrow_{IS-A} \beta$, $\delta \rightarrow_{IS-A} \gamma$, and $W \not\models \neg(\alpha \wedge \delta)$,
   then $(\alpha \; causes \; \beta) \rightarrow \alpha \; explains \; \gamma \; because\_possible \; \{\alpha, \delta\}$

   (b) *Transitivity (gathering the conditions)* If $W \not\models \neg \bigwedge(\Phi \cup \Psi)$, then
   $(\alpha \; explains \; \beta \; because\_possible \; \Phi \wedge \beta \; explains \; \gamma \; because\_possible \; \Psi)$
   $\rightarrow \alpha \; explains \; \gamma \; because\_possible \; (\Phi \cup \Psi)$.

   (c) *Simplification of the set of conditions* If $W \models \bigwedge \Phi \rightarrow \bigvee_{i=1}^{n} \bigwedge \Phi_i$,
   then $\bigwedge_{i \in \{1,\cdots,n\}} \alpha \; explains \; \beta \; because\_possible \; (\Phi_i \cup \Phi)$
   $\rightarrow \alpha \; explains \; \beta \; because\_possible \; \Phi$.

These schemes allow us to obtain the inference patterns described in the previous section:
The base case § 2.2.1: apply (2c) upon (3a) where $\beta = \gamma = \delta$ prior to simplifying by means of (3c).
The upward case § 2.2.2: apply (2c) upon (3a) where $\beta = \delta$, prior to using (3c).
The downward case § 2.2.3: apply (2c) upon (3a) where $\delta = \gamma$.

A more substantial application is: $C = \{\alpha \; causes \; \beta, \; \gamma \; causes \; \epsilon\}$,
$O = \{\beta \rightarrow_{IS-A} \gamma\}$, $W = \{\alpha \rightarrow \beta, \; \beta \rightarrow \gamma, \; \gamma \rightarrow \epsilon\}$
The first form of transitivity in Subsection 2.2.4 requires that we infer:

$$\alpha \; explains \; \epsilon \; because\_possible \; \{\alpha\}$$

Let us proceed step by step:

$\alpha \; explains \; \gamma \; because\_possible \; \{\alpha\}$	by (3a with $\beta = \delta$) as upward case
$\gamma \; explains \; \epsilon \; because\_possible \; \{\gamma\}$	by (3a) as base case
$\alpha \; explains \; \epsilon \; because\_possible \; \{\alpha, \gamma\}$	by (3b)
$\alpha \; explains \; \epsilon \; because\_possible \; \{\alpha\}$	by (3c) simplifying the proviso.

2.4 A generic diagram

Below an abstract diagram is depicted that summarizes many patterns of inferred explanations from various cases of causal statements and $\rightarrow_{IS-A}$ links. The theory is described as follows (see Figure 1):

α *causes* β, α *causes* β_0, β_2 *causes* γ, β_1 *causes* γ,
β_3 *causes* ϵ, γ_1 *causes* δ, γ_3 *causes* δ, ϵ_3 *causes* γ_3;
$\beta \rightarrow_{IS-A} \beta_2$, $\beta_1 \rightarrow_{IS-A} \beta$, $\beta_3 \rightarrow_{IS-A} \beta_0$, $\beta_3 \rightarrow_{IS-A} \beta_1$,
$\gamma_1 \rightarrow_{IS-A} \gamma$, $\gamma_2 \rightarrow_{IS-A} \gamma$, $\gamma_2 \rightarrow_{IS-A} \gamma_3$, $\gamma_2 \rightarrow_{IS-A} \epsilon$,
$\epsilon_1 \rightarrow_{IS-A} \epsilon$, $\epsilon_2 \rightarrow_{IS-A} \epsilon$, $\epsilon_1 \rightarrow_{IS-A} \epsilon_3$, $\epsilon_2 \rightarrow_{IS-A} \epsilon_3$.

This example shows various different "explaining paths" from a few given causal and ontological atoms. Here there is a first "explaining path" from α to δ

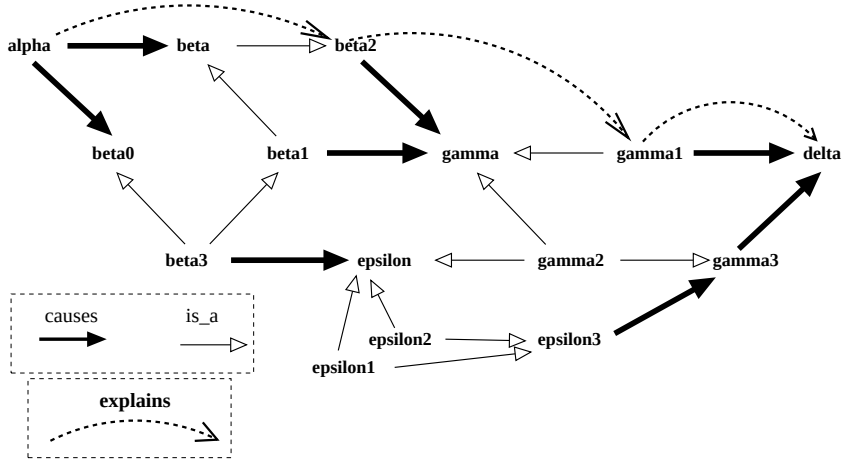

Figure 1: A generic diagram, the theory with a first explaining path

(Figure 1, see also path (1a) on Figure 2). We get successively:α *explains* β_2 *bec_poss* $\{\alpha\}$,

α *explains* γ_1 *bec_poss* $\{\alpha, \gamma_1\}$, and α *explains* δ *bec_poss* $\{\alpha, \gamma_1\}$.
As another "explaining path", we get: α *explains* β_1 *bec_poss* $\{\alpha, \beta_1\}$
α *explains* γ_1 *bec_poss* $\{\alpha, \beta_1, \gamma_1\}$, and α *explains* δ *bec_poss* $\{\alpha, \beta_1, \gamma_1\}$.

This second path is clearly not "optimal" since $\{\alpha, \gamma_1\}$ is strictly included in $\{\alpha, \beta_1, \gamma_1\}$. The simplifying rule produces α *explains* δ *bec_poss* $\{\alpha, \gamma_1\}$ from α *explains* δ *bec_poss* $\{\alpha, \beta_1, \gamma_1\}$ but, from a computational point of view, it is better not to generate the second path at all.

Here there are the four "optimal" explanation atoms from α to δ (see Figure 2 for the precise paths):
(1a) α *explains* δ *bec_poss* $\{\alpha, \gamma_1\}$ (1b) α *explains* δ *bec_poss* $\{\alpha, \gamma_2\}$
(2a) α *explains* δ *bec_poss* $\{\alpha, \beta_3, \epsilon_1\}$ (2b) α *explains* δ *bec_poss* $\{\alpha, \beta_3, \epsilon_2\}$.

We have implemented a program in DLV [LPF$^+$06] (an implementation of the Answer Set Programming (known as ASP) formalism [Bar03]) that takes only a few seconds to give all the results $\sigma 1$ *explains* $\sigma 2$ *bec_poss* Φ, for all examples of this kind, including as here when different explanation paths exist (less than one second for the theory depicted in Figures 1 and 2). See [Moi07]

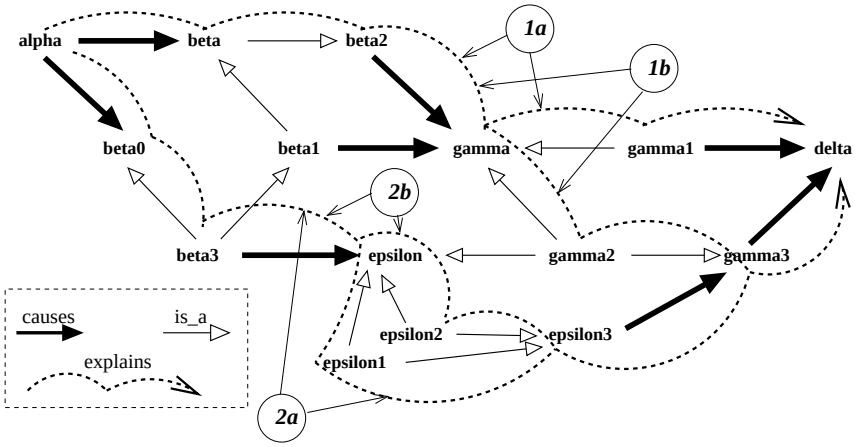

Figure 2: Four optimal explanation paths from "alpha" to "delta"

for details about this implementation. Let us just notice that for now, it does not make full simplifications: it tries to avoid generating explanations which are not simplified, in a way which is optimal in simple cases like this example.

3 Introducing predicates

3.1 Motivations

In order to keep things as simple as possible, we have considered propositional symbols only. However, we have seen various examples where this is not really appropriate. Indeed, stating $Heard\_loud\_bell \rightarrow_{IS-A} Heard\_bell$ is neither natural nor convenient. Indeed, we should also state $Activating\_loud\_bell \rightarrow_{IS-A} Activating\_bell$ if necessary and so on. It is clear that it is much more natural to state $loud\_bell \rightarrow_{IS-A} bell$, and to infer the results about $Heard\_[...]$, $Activating\_[...]$ and so on.

We introduce predicate symbols (such as $Heard$, $Activating$, Own, ...) and constant symbols such as $bell, noise, loud\_bell, student, book, ...$

The elementary terms (represented by α, β as in preceding sections) are ground atoms such as $Heard(bell)$, $On(alarm)$, $Own(student, book)$. We need a way to use the ontological information together with the causal information. The ontological links concern constant symbols, as in e. g. $a \rightarrow_{IS-A} b$.

We need a way to infer our old $Heard\_loud\_bell \rightarrow_{IS-A} Heard\_bell$ from the new ontological information stating here $loud\_bell \rightarrow_{IS-A} bell$. Again, since we want to keep things simple, we put some restrictions. We distinguish two kinds of behavior for a given parameter in a predicate (this is the main improvement from [BCM07]).

Let us suppose that $Heard(bell)$ means "I have heard some bell". Then, we

can say that $Heard$ is ***essentially existential*** (*there exists* some bell that I have heard). A more explicit way to express this is to denote this predicate by $Heard_{one}$ instead of $Heard$. Similarly, let us suppose that $Own(student, book)$ means that "every student owns some book". We will say that the predicates $Heard$ and Own ***inherit upward*** for the parameter t in $Heard(t)$ and in $Own(t1, t)$ through the $IS-A$ hierarchy.

The other particular case of predicates, which ***inherit downward*** [for a given parameter] through the $IS-A$ hierarchy, are ***essentially universal*** [for this parameter]. Let us take $Like$ as an example, considering that $Like(bell)$ means I like bells (in general: I like *all* the bells). This predicate could be denoted by $Like_{all}$. This notation allows to use the two predicates $Like_{one}$ and $Like_{all}$ together, if necessary, and has the advantage of indicating the kind of inheritance of a predicate in its denomination. Similarly, Own inherits downward for the parameter $t1$ in $Own(t1, t)$ and could be denoted by $Own_{all,one}$.

This problem of the way predicates should inherit through the $IS-A$ hierarchy, is a matter of formalization of natural language. The way used here allows to avoid explicit quantifiers such as $\exists$ and $\forall$. When introducing a predicate, we must state [for each of its parameters] whether it is essentially existential (then it inherits upward such as $Heard_{one}$) or essentially universal (then it inherits downward such as $Like_{all}$), if we want to take advantage of the ontological information with respect to this predicate. Predicates which are neither "essentially existential" nor "essentially universal" for some of their parameters cannot exploit the $IS-A$ hierarchy for these parameters.

Let us provide the formal description of the new system.

3.2 The vocabulary with "upward and downward inheriting" predicates

1. ***Classical vocabulary and formulas***

   (a) ***Predicate symbols***. Any arity is possible and, for each of their parameters, predicates can be *essentially existential* or *essentially universal* or without precision. The names of the predicate symbols begin with an uppercase letter such as in $P, Heard, On$.

   (b) ***Constant symbols***. Their names begin with a lower case letter such as in $a, bell, loud\_bell$.

   (c) ***Classical atoms***. A classical atom is a ground atom $P(a_1, ..., a_n)$ where P is a predicate of arity n and a_i's are constants. A propositional symbol (i.e. a predicate of arity 0) P_0 is thus a classical atom. Classical atoms are denoted either as $P(a_1, ..., a_n)$ or $P_1(a)$ or by Greek letters such as α or β_1.

   (d) ***Classical formulas***. A classical formula is a sentence (Boolean combinations of classical atoms: only ground formulas are considered).

2. ***Causal atoms and formulas***

(a) If α and β are classical atoms, then α *causes* β is a ***causal atom***.

(b) A ***causal formula*** is a Boolean combination of classical or causal atoms.

3. ***Ontological atoms***
If a and b are constant symbols, then $a \rightarrow_{IS-A} b$ is an ontological atom. Since we want our "predicate" system to encompass the preceding "propositional" system, we must also consider ontological links between two propositional symbols: if P_0 and Q_0 are two propositional symbols, then $P_0 \rightarrow_{IS-A} Q_0$ is an ontological atom.

4. A ***causal theory*** consists of a set W of classical formulas, a set C of causal formulas and a set O of ontological atoms.

5. ***Explanation atoms*** From a given causal theory, some *explanation atoms* will be derived. An *explanation atom* is

$$\alpha \ \ explains \ \ \beta \ \ because\_possible \ \ \Phi$$

where α and β are classical atoms and Φ is a set of classical atoms. It reads "α explains β because the set Φ is possible".

Basically, the derivation of the explanation atoms is as given in the propositional case, thus now we can introduce directly the ***formal proof system***.

3.3 Formal proof system of the formalism with predicates

1. ***Property of the causal atoms: entailing implication***

(a) $(\alpha \ causes \ \beta) \rightarrow (\alpha \rightarrow \beta)$

2. ***Ontological atoms***
The easiest way to present the rules in the predicate case is to augment the ontology by introducing ontological links between ground atoms:

(a) ***Deriving an augmented ontological relation***

i. If $\alpha = P_0$ and $\beta = Q_0$ are propositional atoms, then $P_0 \rightarrow_{IS-A(aug.)} Q_0$ whenever $P_0 \rightarrow_{IS-A} Q_0$.

ii. Let $P_{all,one,\_,\cdots}$ denote some predicate of arity n, essentially universal with respect to its first parameter and essentially existential with respect to its second parameter (other parameters not concerned here, clearly the first or the second parameter could similarly by the i^{th} parameter for any $i \in \{1, \cdots, n\}$). Then $P_{all,one,\_,\cdots}(a_1, b_2, a_3, \cdots, a_n) \rightarrow_{IS-A(aug.)} P_{one}(a_1, a_2, a_3, \cdots, a_n)$ whenever $a_2 \rightarrow_{IS-A} b_2$, and $P_{all,one,\_,\cdots}(a_1, a_2, a_3, \cdots, a_n) \rightarrow_{IS-A(aug.)} P_{one}(b_1, a_2, a_3, \cdots, a_n)$ whenever $a_1 \rightarrow_{IS-A} b_1$.

(b) ***Properties of the augmented ontological relation***

i. ***Transitivity***
If $\alpha \rightarrow_{IS-A(aug.)} \beta$ and $\beta \rightarrow_{IS-A(aug.)} \gamma$ then $\alpha \rightarrow_{IS-A(aug.)} \gamma$.

ii. ***Reflexivity*** $\alpha \rightarrow_{IS-A(aug.)} \alpha$.

iii. ***Entailing implication*** If $\alpha \rightarrow_{IS-A(aug.)} \beta$ then $\alpha \rightarrow \beta$.

The only difference with the propositional case is that we must use the augmented ontology instead of the ontology given by the user.

3. ***Deriving explanation atoms***

(a) ***Base case***
If $\beta \rightarrow_{IS-A(aug.)} \gamma$, $\beta \rightarrow_{IS-A(aug.)} \delta$, and $W \not\models \neg(\alpha \wedge \beta)$
then α *causes* $\gamma \rightarrow \alpha$ *explains* δ *because_possible* $\{\alpha, \beta\}$.

(b) ***Transitivity of explanation*** cf Point 3b in § 2.3.

(c) ***Simplifying explanation atoms*** cf Point 3c in § 2.3.

Notice that we keep transitivity and reflexivity of the ontology, for the augmented relation $\rightarrow_{IS-A(aug.)}$. As for the ontology relation $\rightarrow_{IS-A}$, which is the relation provided by the user, we could add these two properties if desired: this would not modify the explanation atoms.

Here there is an example, illustrating also the interest of the general generating rule of explanations, with "down then up" ontological links (Point 3a):

$C1$ $On(alarm)$ *causes* $Heard(warning\_signal)$;
$C2$ $Heard(loud\_noise)$ *causes* $Wake\_up$;
$O1$ $loud\_bell \rightarrow_{IS-A} warning\_signal$,
$O2$ $hooter \rightarrow_{IS-A} warning\_signal$,
$O3$ $loud\_bell \rightarrow_{IS-A} loud\_noise$,
$O4$ $red\_flashing\_light \rightarrow_{IS-A} warning\_signal$.

$Heard$ is supposed to be essentially existential, thus it inherits upward (for On, it does not matter in this example).

We get the following "augmented" ontological links:

$O1aug$: $Heard(loud\_bell) \rightarrow_{IS-A(aug.)} Heard(warning\_signal)$,
$O2aug$: $Heard(hooter) \rightarrow_{IS-A(aug.)} Heard(warning\_signal)$,
$O3aug$: $Heard(loud\_bell) \rightarrow_{IS-A(aug.)} Heard(loud\_noise)$,
$O4aug$: $Heard(red\_flashing\_light) \rightarrow_{IS-A(aug.)} Heard(warning\_signal)$.

Thus, we get the following explanation atoms:

$E1$: $On(alarm)$ *explains* $Heard(loud\_noise)$
bec_poss $\{On(alarm), Heard(loud\_bell)\}$
$E2$: $Heard(loud\_noise)$ *explains* $Wake\_up$ *bec_poss* $\{Heard(loud\_noise)\}$
$E1$ comes from $C1, O1aug$ and $O3aug$, and
$E2$ from $C2$, by the base case of explanations.

Then we get, by transitivity of explanations on $E1$ and $E2$,

$On(alarm)$ *explains* $Wake\_up$ *bec_possible* $\{On(alarm),$
$Heard(loud\_bell), Heard(loud\_noise)\}$ and finally
$On(alarm)$ *explains* $Wake\_up$ *bec_poss* $\{On(alarm), Heard(loud\_bell)\}$
by simplifying the set of conditions, taking into account that we get
$Heard(loud\_bell) \rightarrow Heard(loud\_noise)$ from $O3aug$.

As a formal example, let us take the predicate variant of the example ending § 2.3, where P denotes a unary predicate which is essentially existential for its parameter and γ some arbitrary classical atom.

$C = \{P(a)$ *causes* $P(b)$, $P(c)$ *causes* $\gamma\}$,
$O = \{b \rightarrow_{IS-A} c\}$,
$W = \{P(a) \rightarrow P(b),\ \ P(b) \rightarrow P(c),\ \ P(c) \rightarrow \gamma\}$
As in the example ending § 2.3, we get:

$$P(a) \textit{ explains } \gamma \textit{ because\_possible } \{P(a)\}.$$

Again, let us proceed step by step:

$P(b) \rightarrow_{IS-A(aug.)} P(c)$	by (2(a)ii)
$P(a)$ *explains* $P(c)$ *because_possible* $\{P(a)\}$	by (3a) as upward case
$P(c)$ *explains* γ *because_possible* $\{P(c)\}$	by (3a) as base case
$P(a)$ *explains* γ *because_possible* $\{P(a), P(c)\}$	by (3b)
$P(a)$ *explains* γ *because_possible* $\{P(a)\}$	by (3c) simplifying the proviso

As an example of a predicate of arity greater than 1, let us define a binary predicate Own where

$Own(student, book)$ is intended to mean "every student owns a book".

Let us suppose that our ontology contains $mary \rightarrow_{IS-A} student$,
$student \rightarrow_{IS-A} human$ and $book \rightarrow_{IS-A} written\_document$.

Notice that we allow "reification" in our formalism: concepts such as "student", "human" and "book" are represented by constants, exactly as are "individuals" such as "Mary". Since Own is intended to mean here "owns a" and not "owns all", this binary predicate is essentially existential (thus it inherits upward) with respect to its second parameter. The case of the first parameter has been settled also since here $Own(student, book)$ means "every student owns a book": Own is essentially universal with respect to its first parameter. If we need also another predicate Own' where $Own'(student, book)$ means *there exists a student owning a book*, it is more convenient to denote Own by $Own_{all,one}$ and Own' by $Own_{one,one}$.

We must state explicitly whether a predicate is essentially existential ("*one*" kind) or essentially universal ("*all*" kind) (or none of these two options by default), with respect to each of its parameters. This kind of problem occurs each time we want to formalize natural language: the user must be aware that it is important to make the intended meaning of each predicate precise.

It is an interesting feature of our formalism that this precise meaning can be expressed in a natural way (at least if this predicate can be used from ontological

atoms). So, it is not enough to mention the arity of a predicate, its "*one*" or "*all*" kind should be given for each of its parameters.

This will indicate to the system, for each parameter of a predicate, whether the inheritance with respect to the ontology is "upward" ("*one*" kind parameter) or "downward" ("*all*" kind parameter). It is possible to use parameters for which neither the "*one*" kind nor the "*all*" kind applies. Let us consider such a predicate $P_{all,na,one}$ ("na" for "not available"). Then, no augmented ontological link exists between atoms of this predicate where this second parameter has different values on the left side and on the right side: if $P(t_1, t_2, t_3) \rightarrow_{IS-A(aug.)} P(t'_1, t'_2, t'_3)$ is produced, then $t_2 = t'_2$.

We would get here:
$Own_{all,one}(human, book) \rightarrow_{IS-A(aug.)} Own_{all,one}(mary, book)$,
$Own_{all,one}(human, written\_document) \rightarrow_{IS-A(aug.)} Own_{all,one}(mary, written\_document)$,
$Own_{all,one}(mary, book) \rightarrow_{IS-A(aug.)} Own_{all,one}(mary, written\_document)$,
$Own_{all,one}(human, book) \rightarrow_{IS-A(aug.)} Own_{all,one}(human, written\_document)$.

Thus, we would also get by transitivity of $\rightarrow_{IS-A(aug.)}$:
$Own_{all,one}(human, book) \rightarrow_{IS-A(aug.)} Own_{all,one}(mary, written\_document)$.

3.4 Extending the formalism: ontological links between predicates

We could even extend the formalism so that it allows ontological links between predicates of arity 1 or more. As an example, let us suppose that we have the two unary predicates (of the "one" kind, but it does not really matter here) $Heard$ and $Perceived$. It is natural to state $Heard \rightarrow_{IS-A} Perceived$.

We would add the following to Point 2 § 3.3 (and similarly for higher arities):

($2_{ext}{}'$) ***Ontological atoms: introducing an augmented relation from an ontology between predicates***
If P, Q are unary predicates, then
if $P \rightarrow_{IS-A} Q$ then $P(a) \rightarrow_{IS-A(aug.)} Q(a)$.

In our example, we would get $Heard(bell) \rightarrow_{IS-A(aug.)} Perceived(bell)$, $Heard(noise) \rightarrow_{IS-A(aug.)} Perceived(noise)$.

In this way, it is easier and more natural to express various relations between "events". Notice that these general ontological links between predicate symbols generalize in a natural way the ontological links given above (Point 2(a)i in § 3.3) for propositional symbols.

It can be noticed that, with the ASP translation evoked above, predicates and constants are represented by ASP constants anyway, thus it is not harder to include the ontology between predicates.

Here there is the last example of what can be expected from this formalism:

Example. Getting cold usually causes Mary to become active. I see Mary jogging. So, Mary getting cold might be taken as an explanation for her jogging.

$C = \{Getting\_cold(mary)\ causes\ Moving\_up(mary)\ \}$

$O = \{Jogging \rightarrow_{IS-A} Moving\_up\}$.

For now, W does not contain any special information (only the consequences of the preceding causal and ontological atoms).

We get $Jogging(Mary) \rightarrow_{IS-A(aug.)} Moving\_up(Mary)$ from $(2_{ext}{}')$.

Thus $W = \{\ Getting\_cold(mary) \rightarrow Moving\_up(mary),$
$Jogging(mary) \rightarrow Moving\_up(mary).\}$

"Mary getting cold" can be inferred as an explanation for "Mary is jogging". Indeed, the causal theory entails $Getting\_cold(mary) explains\ Jogging(mary)$
$bec\_poss\{Getting\_cold(mary), Jogging(mary)\}$. $(EXPL)$

If now we add the fact that if the weather is not cold, then Mary cannot get cold, this explanation is no longer possible in warm weather.

Adding the following formulas to W takes the new information into account:
$Warm\_Weather$, $\neg(Warm\_Weather \wedge Cold\_Weather)$,
$\neg Cold\_Weather \rightarrow \neg Getting\_cold(mary)$.

Then, the causal theory fails to entail the explanation atom $(EXPL)$.

4 About a few features of the formalism

The explanation inference follows the patterns presented before. The inference pattern (3a) in § 2.3 (or pattern (3a) in § 3.3) is important.

1. In this inference pattern, the direction of the $\rightarrow_{IS-A}$ links $a \rightarrow_{IS-A} b$ and $a \rightarrow_{IS-A} c$ is important. Unexpected conclusions would ensue if other directions, e.g., $b \rightarrow_{IS-A} a$ and $c \rightarrow_{IS-A} a$, were allowed.

2. Also, there are good reasons for excluding conclusions not endorsed by this pattern. One reason is that it is better to limit the explanations to a minimum, otherwise an overwhelming set of "explanations" could result. As an example of conclusions not endorsed, notice that no explanation atom can be derived if it does not start with a ground atom which occurs somewhere on the left side of a causal atom.

As a short justification of these two points, let us introduce an example which comes from a real-world application [SMOC$^+$98, BC99], but has been drastically simplified and reduced. The causal model describes a physical system in which a *sliding of the flywheel* (to be abbreviated as *SOF*) causes a *step* in the vibration measurement signal. It is also known that *step* and *slow increase* are two kinds of evolution of the vibration measurement signal and that a *sharp step* is itself a kind of step.

This is to be formalized as follows (the example has been simplified for the sake of conciseness and clarity, thus the propositional version of § 2 suffices).

(C1) SOF *causes* $Step$;
(O1) $Step \rightarrow_{IS-A} Evolution$, (O2) $Slow\_increase \rightarrow_{IS-A} Evolution$
(O3) $Sharp\_step \rightarrow_{IS-A} Step$.
Here are the explanation atoms which can be derived:
($E1$) SOF *explains* $Step$ *bec_poss*$\{SOF\}$,
($E2$) SOF *explains* $Evolution$ *bec_poss*$\{SOF\}$,
($E3$) SOF *explains* $Sharp\_step$ *bec_poss*$\{SOF, Sharp\_step\}$.

Let us consider Point 1 above: Our concern here is about
SOF *explains* $Sharp\_step$ *bec_poss*$\{SOF, Sharp\_step\}$ which can be inferred versus
SOF *explains* $Slow\_increase$ *bec_poss*$\{SOF, Slow\_increase\}$ which cannot.

Why is it sensible to explain "evolution" and "sharp step" by a "sliding of the flywheel" while "slow increase" could not be explained by the same "sliding of the flywheel"? The reason is that, from the facts given here, a "slow increase" is not a "step" (which is indeed explained by some "sliding of the flywheel"), but another kind of "evolution". So, explaining "slow increase" by some "sliding of the flywheel" would be unmotivated from what we know about the system. On the other hand, it is possible that some "sharp step", which is a kind of "step", has been provoked by some "sliding of the flywheel". Notice that if there were reasons to eliminate this possibility, it should have been noticed. We could e.g. have added the information $\neg(SOF \wedge Sharp\_step)$ in W. Another way would be to replace $C1$ by ($C1'$) SOF *causes* $Moderate\_step$,
and to modify the ontological information accordingly.

This example also shows that we must exercise some care while stating the ontological information, which was to be expected since the ontological information plays a great rôle in the formalism. Since a "step" is an "evolution", there are good reasons to explain "evolution" also by some "sliding of the flywheel": If we have enough information to know that we get "evolution", but not enough to know whether it is a "step" or not, it is natural to provide "SOF" as an explanation for this "evolution".

Let us consider Point 2 now. For this purpose, let us add the following information about the system: "Any *evolution* in the vibration measurement signal causes an *alarm* to be displayed on the operator control screen."

We would then add the following formula to our causal theory.

($C2$) $Evolution$ *causes* $Alarm$.

The theory is now described by $C1$, $C2$ together with $O1$, $O2$ and $O3$ (W does not contain special information, only the consequences of these formulas).

Then, the explanation atoms derived by the new theory are $E1$, $E2$, $E3$ and $E4$, plus $E2'$ obtained by transitivity from $E2$ and $E4$ and by an obvious simplification, with $E4$ and $E2'$ as follows:
($E4$) $Evolution$ *explains* $Alarm$ *bec_poss*$\{Evolution\}$,
($E2'$) SOF *explains* $Alarm$ *bec_poss*$\{SOF\}$.

We do not get $(notE2)$ $Step$ $explains$ $Alarm$ $bec\_poss\{Step\}$.

The reason is that we restrict our "explanations" to those starting from a ground atom which actually "causes" something, and $Step$ does not appear on the left side of a *causal atom*. Notice however that, if $Step$ is established, then so is $Evolution$ from $O1$, thus, from $E4$ we indeed get some "explanation" in which the left-hand side is established while the right-hand side is $Alarm$.

It must be noticed that here are cases (particularly when making abduction from a set of atoms) where it is convenient to derive also "explanations" starting from atoms such as $Step$ here. This addition is immediate in our formalism.

5 Conclusion

We have provided a logical framework allowing predictive and abductive reasoning from *causal* information. Indeed, the formalism allows to express causal information in a direct way. Then, we deduce so-called *explanation atoms* which capture what might explain what, in view of the given information. We have resorted to *ontological* information, which is key in generating sensible explanations from causal statements.

The user provides taxonomic information as a list of *ontological atoms* $a \rightarrow_{IS-A} b$ intended to mean that object a "is a" b. The basic ground atoms, denoted α, β, are then built with predicates, such as $P(a, b)$. The user provides causal information as causal atoms α $causes$ β (which can occur in more complex formulas). This makes formalization fairly short and natural. The ontology is used in various patterns of inference for explanations. Such information is easy to express, or to obtain in practice, due to existing ontologies and ontological languages. If we were in a purely propositional setting, the user should write $Own\_small\_car \rightarrow_{IS-A} Own\_car$, $Own\_big\_car \rightarrow_{IS-A} Own\_car$, and also $Heard\_small\_car \rightarrow_{IS-A} Heard\_car$ and so on.

This would be cumbersome. In contrast, our setting is "essentially propositional" for what concerns the causal atoms, in that it is as if $Own(small\_car)$ were a propositional symbol Own_small_car, while, for what concerns the ontology, we really use the fact that $Heard$ and Own are predicates.

The notion of predicates "essentially existential" or "universal" allows to keep a "datalog" formalism, very closed to a propositional one, without the need for explicit quantifiers $\forall$ *(for all)* or $\exists$ *(there exists)*.

The present proposal is a compromise between simplicity, as well as clarity, when it comes to describing a situation, and efficiency and pertinence of the results provided by the formalism.

Our work differs from other approaches in the literature in that it strictly separates causality, ontology and explanations. The main advantages are that information is more properly expressed and that our approach is compatible with various accounts of these notions, most notably causality. In particular,

we need no special instances of α *causes* α to hold (even though α *causes* α for a particular α can be explicitly asserted). Similarly, if α is equivalent to γ and β *causes* α hold, this does not mean that β *causes* γ holds. This feature contrasts with [Bel06, Boc03, GLL$^+$04, HP01a, HP01b, G.98] although in the context of actions such confusion is less harmful. Some authors have already introduced notions related to our causal and ontological atoms. In [Kau91, CD94], there are "axioms" which can be loosely related to our causal and ontological atoms. Our work investigates how "explanations" are obtained from such causal and ontological information. On the other hand, we have not worked here on the important subject of what can precisely be done from these explanation atoms. This is left for future work, since we think that the formalization task is also a crucial one. As for designing some plan recognition or some abductive reasoning from our work, this is possible with simple additions over our formalism. Let us just give one indication here: from the sets of conditions for each explanation atom, "best explanations" for a given set set of α's can be defined.

Also as future work we should relax some of the strong restrictions on the notion of "cause" made here. We could introduce some ranking among the causal atoms, in order to cope with cases such as "smoking causes cancer". Then, we could introduce the temporal aspect which is important as soon as causation is involved, by adding a special temporal parameter.

We have designed a system in answer set programming that implements most of the formalism introduced above. It is restricted to predicates of arities 0 or 1 (this could be easily extended) and the simplification part is not fully completed (this is much harder to modify, since the computation would be seriously more complex, but this strong simplification does not appear to be crucial). It works for examples of reasonable size, but it should be expandable for some real life examples, thus showing that our two main goals have been reached (simplicity of the formalization by a user, and efficiency of the computation). Our present translation in Answer set programming cannot be considered as a competitor with achieved proposals such as the "causal calculator" CCalc of [GLL$^+$04], but it can at least give some indications that a real abductive system (for instance) can be built over our present proposal. We can e.g. deal with the "cooking example" of [Kau91, CD94] by adding only a few rules to our program.

As another future work, we should consider ontological links less elementary than the taxonomic relations considered in the present system. We think however that the present system is a good basis for a really practical system.

References

[Bar03] Chitta Baral. *Knowledge representation, reasoning and declarative problem solving.* Cambridge University Press, January 2003, 2003.

[BC99] Ph. Besnard and M.-O. Cordier. Inferring causal explanations. In *5th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning, in LNAI 1638, Springer*, pages 55–67, London, 1999.

[BCM06] Philippe Besnard, Marie-Odile Cordier, and Yves Moinard. Configurations for Inference between Causal Statements. In Jrme Lang, Fangzhen Lin, and Ju Wang, editors, *Knowledge Science, Engineering and management, in LNAI 4092*, pages 292–304, Guilin, China, August 2006. Springer.

[BCM07] Philippe Besnard, Marie-Odile Cordier, and Yves Moinard. Ontology-based inference for causal explanation. In Zili Zhang and Jrg Siekmann , editor, *Knowledge Science, Engineering and management, in LNAI 4798*, pages 153–164, Melbourne, Australia, November 2007. Springer.

[Bel06] J. Bell. Causation as Production. In G. Brewka, S. Coradeschi, A. Perini, and P. Traverso, editors, *17th biennial European Conference on Artificial Intelligence*, pages 327–331, Riva del Garda, Italy, 2006. IOS Press.

[Boc03] Alexander Bochman. A Logic for Causal Reasoning. In Georg Gottlob and Toby Walsh, editors, *IJCAI-03*, pages 141–146, Acapulco, Mexico, August 2003. Morgan Kaufmann.

[CD94] L. Console and D. Theseider Dupré. Abductive reasoning with abstraction axioms. *Lecture Notes in Computer Science*, 810:98–112, 1994.

[G.98] Shafer G. Causal Logic. In H. Prade, editor, *ECAI-98*, pages 711–720, Brighton, UK, 1998. Wiley.

[GLL+04] Enrico Giunchiglia, Joohyung Lee, Vladimir Lifschitz, Norman McCain, and Hudson Turner. Nonmonotonic causal theories. *Artficial Intelligence*, 153(1–2):49–104, March 2004.

[HP01a] J. Halpern and J. Pearl. Causes and Explanations: A Structural-Model Approach. Part I: Causes. In J. S. Breese and D. Koller, editors, *UAI-01*, pages 194–202, Seattle, 2001. Morgan Kaufmann.

[HP01b] J. Halpern and J. Pearl. Causes and Explanations: A Structural-Model Approach. Part II: Explanations. In J. S. Breese and D. Koller, editors, *IJCAI-01*, pages 27–34, Seattle, 2001. Morgan Kaufmann.

[Kau91] H. A. Kautz. A formal theory of plan recognition and its implementation. In J. F. Allen, H. A. Kautz, R. Pelavin, and J. Tenenberg, editors, *Reasoning About Plans*, pages 69–125. Morgan Kaufmann Publishers, San Mateo (CA), USA, 1991.

[LPF+06] Nicola Leone, Gerald Pfeifer, Wolfgang Faber, Thomas Eiter, Georg Gottlob, Simona Perri, and Francesco Scarcello. The DLV System for Knowledge Representation and Reasoning. *ACM Transactions on Computational Logic (TOCL)*, 7(3):499–562, 2006.

[Moi07] Yves Moinard. An Experience of Using ASP for Toy Examples. In Stefania Costantini and Richard Watson, editors, *Advances in Theory and Implementation*, pages 133–147, Porto, Portugal, September 8 and 13 2007. Facultade de Ciencias, Universidade do Porto.

[SMOC+98] Cauvin S., Cordier M.-O., Dousson C., Laborie P., Lévy F., Montmain J., Porcheron M., Servet I. and Travé-Massuyès L. Monitoring and alarm interpretation in industrial environments. *AI Communications*, 11(3–4):139–173, 1998.

Acknowledgement

The authors thank the reviewers for their helpful and constructive comments.