Splitting Epistemic Logic Programs

Pedro Cabalar
University of Corunna
Corunna, Spain
cabalar@udc.es

Jorge Fandinno
University of Toulouse
IRIT, CNRS, France
jorge.fandinno@irit.fr

Luis Fariñas del Cerro
University of Toulouse
IRIT, CNRS, France
farinas@irit.fr

Abstract

Epistemic logic programs constitute an extension of the stable models semantics to deal with new constructs called *subjective literals*. Informally speaking, a subjective literal allows checking whether some regular literal is true in all stable models or in some stable model. As it can be imagined, the associated semantics has proved to be non-trivial, as the truth of the subjective literal may interfere with the set of stable models it is supposed to query. As a consequence, no clear agreement has been reached and different semantic proposals have been made in the literature. Unfortunately, comparison among these proposals has been limited to a study of their effect on individual examples, rather than identifying general properties to be checked. In this paper, we propose an extension of the well-known splitting property for logic programs to the epistemic case. To this aim, we formally define when an arbitrary semantics satisfies the epistemic splitting property and examine some of the consequences that can be derived from that, including its relation to conformant planning and to epistemic constraints. Interestingly, we prove (through counterexamples) that most of the existing proposals fail to fulfill the epistemic splitting property, except the original semantics proposed by Gelfond in 1991.

Introduction

The language of *epistemic specifications*, proposed by Gelfond (1991), constituted an extension of disjunctive logic programming that introduced modal operators to quantify over the set of stable models (Gelfond and Lifschitz, 1988) of a program. These new constructs were later incorporated as an extension of the Answer Set Programming (ASP) paradigm in different solvers and implementations – see (Leclerc and Kahl, 2018b) for a recent survey. The new constructs, *subjective literals*, have the form $\mathbf{K}\,l$ and $\mathbf{M}\,l$ and allow respectively checking whether regular literal l is true in every stable model (cautious consequence) or in some stable model (brave consequence) of the program. In many cases, these subjective literals can be seen as simple queries, but what makes them really interesting is their use in rule bodies, what may obviously affect the set of stable models they are meant to quantify. This feature makes them suitable for modelling introspection but, at the same time, easily involves cyclic specifications whose intuitive behaviour is not always easy to define. For instance, the semantics of an epistemic logic program may yield alternative sets of stable models, each set being called a *world view*. Deciding the intuitive world views of a cyclic specification has motivated a wide debate in the literature. In fact, in (Gelfond, 1991) original semantics or in its extension (Truszczyński, 2011), some cyclic examples manifested self-supportedness, so Gelfond (2011) himself and, later on, other authors (Kahl, 2014; Kahl et al., 2015; Fariñas del Cerro, Herzig, and Su, 2015; Shen and Eiter, 2017; Son et al., 2017) proposed different variants trying to avoid unintended results, without reaching a clear agreement. Unfortunately, comparison among these variants was limited to studying their effect on a set of "test" examples. This methodology has proven to fall short in such an uncertain context: confidence in any proposal is always subject to the appearance of new counterintuitive examples. A much stronger method would be defining instead formal properties to be established, as this would cover complete families of examples and, hopefully, could help to reach an agreement on some language fragments. For instance, one would expect that, at least, the existing approaches agreed on their interpretation of acyclic specifications. Regretfully, as we will show later, this is not the case.

In this paper we propose a candidate property, we call *epistemic splitting*, that not only defines an intuitive behaviour for stratified epistemic specifications but also goes further, extending the splitting theorem (Lifschitz and Turner, 1994), well-known for standard logic programs, to the epistemic case. Informally speaking, we say that an epistemic logic program can be split if a part of the program (the *top*) only refers to the atoms of the other part (the *bottom*) through subjective literals. A given semantics satisfies epistemic splitting if, given any splitted program, it is possible to get its world views by first obtaining the world views of the bottom and then using the subjective literals in the top as a "query" on the bottom part previously ob-

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

tained. If epistemic splitting holds, the semantics immediately satisfies other properties. For instance, if the use of epistemic operators is stratified, the program has a unique world view at most. Similarly, epistemic constraints (those only consisting of subjective literals) can be guaranteed to be monotonic: they only rule out candidate world views. As we will see, however, only (Gelfond, 1991) satisfies epistemic splitting among the previously cited approaches. So, somehow, the recent attempts to fix the behaviour of cycles has neglected the attention on the effects produced on acyclic specifications. In fact, a different property of epistemic splitting was already proved by (Watson, 2000) as a method to compute world views for (Gelfond, 1991). However, this definition is based on a "safety" condition that needs to be checked for all possible world views and is specific for (Gelfond, 1991) semantics, so it is harder to justify as a general property required for other approaches.

The rest of the paper is organized as follows. First, we motivate the main idea through a well-known example. After that, we recall basic definitions of (non-epistemic) answer set programming and splitting, introduce the language of epistemic specifications and define (Gelfond, 1991) semantics. In the next section, we proceed to define the property of epistemic splitting and study some of its consequences. Then, we formally prove that (Gelfond, 1991) satisfies this property while we provide counterexamples for the other approaches, concluding the paper after that.

Motivation

To illustrate the intuition behind our proposal, let us consider the well-known standard example introduced in (Gelfond, 1991).

Example 1. *A given college uses the following set of rules to decide whether a student X is eligible for a scholarship:*

$$\mathit{eligible}(X) \leftarrow \mathit{high}(X) \quad (1)$$
$$\mathit{eligible}(X) \leftarrow \mathit{minority}(X), \mathit{fair}(X) \quad (2)$$
$$\sim\mathit{eligible}(X) \leftarrow \sim\mathit{fair}(X), \sim\mathit{high}(X) \quad (3)$$

Here, '$\sim$' stands for strong negation and $\mathit{high}(X)$ and $\mathit{fair}(X)$ refer to the grades level of student X. We want to encode the additional college criterion:

> *"The students whose eligibility is not determined by the college rules should be interviewed by the scholarship committee."*

as another rule in the program. □

The problem here is that, for deciding whether $\mathit{eligible}(X)$ "*can be determined*," we need to check if it holds in all the answer sets of the program, that is, if it is one of the cautious consequences of the latter. For instance, if the only available information for some student *mike* is the disjunction

$$\mathit{fair}(\mathit{mike}) \vee \mathit{high}(\mathit{mike}) \quad (4)$$

we get that program $\{(1)-(4)\}$ has two stable models, $\{\mathit{fair}(\mathit{mike})\}$ and $\{\mathit{high}(\mathit{mike}), \mathit{eligible}(\mathit{mike})\}$ so $\mathit{eligible}(\mathit{mike})$ cannot be determined and an interview should follow. Of course, if we just want to query cautious and brave consequences of the program, we can do it inside ASP. For instance, the addition of constraint:

$$\bot \leftarrow \mathit{eligible}(\mathit{mike})$$

allows us deciding if $\mathit{eligible}(\mathit{mike})$ is a cautious consequence by just checking that the resulting program has no answer sets. The difficulty comes from the need to *derive* new information from a cautious consequence. This is where subjective literals come into play. Rule

$$\mathit{interview}(X) \leftarrow \texttt{not}\,\mathbf{K}\,\mathit{eligible}(X), \texttt{not}\,\mathbf{K}\,\sim\mathit{eligible}(X) \quad (5)$$

allows us to prove that $\mathit{interview}(X)$ holds when neither $\mathit{eligible}(X)$ nor $\sim\mathit{eligible}(X)$ are cautious consequences of (1)-(4). The novel feature here is that (5) is also part of the program, and so, it affects the answer sets queried by $\mathbf{K}$ too, which would actually be:

$$\{\mathit{fair}(\mathit{mike}), \mathit{interview}(\mathit{mike})\} \quad (6)$$
$$\{\mathit{high}(\mathit{mike}), \mathit{eligible}(\mathit{mike}), \mathit{interview}(\mathit{mike})\} \quad (7)$$

So, there is a kind of cyclic reasoning: operators $\mathbf{K}$ and $\mathbf{M}$ are used to query a set of stable models that, in their turn, may depend on the application of that query. In the general case, this cyclic reasoning is solved by resorting to multiple world views, but in our particular example, however, this does not seem to be needed. One would expect that separating the queried part (1)-(4) and the rule that makes the query (5) should be correct, since the first four rules do not depend on (5) and the latter exclusively consults them without interacting with their results. This same reasoning could be applied if we added one more level such as, for instance, by including the rule:

$$\mathit{appointment}(X) \leftarrow \mathbf{K}\,\mathit{interview}(X) \quad (8)$$

The two answer sets of program $\{(1)-(5)\}$ contain $\mathit{interview}(\mathit{mike})$ and so $\mathit{appointment}(\mathit{mike})$ can be added to both answer sets incrementally. This method of analysing a program by division into independent parts shows a strong resemblance to the *splitting theorem* (Lifschitz and Turner, 1994), well-known in standard ASP. Splitting is applicable when the program can be divided into two parts, the *bottom* and the *top*, in such a way that the bottom never refers to head atoms in the top. When this happens, we can first compute the answer sets of the bottom and then, for each one, simplify the top accordingly, getting new answer sets that complete the information. We could think about different ways of extending this method for the case of epistemic logic programs, depending on how restrictive we want to be on the programs where it will be applicable. However, we will choose a very conservative case, looking for a wider agreement on the proposed behaviour. The condition we will impose is that our top

program can only refer to atoms in the bottom through epistemic operators. In this way, the top is seen as a set of rules that derive facts from epistemic queries on the bottom. Thus, each world view W of the bottom will be used to replace the subjective literals in the top by their truth value with respect to W. For the sake of completeness, we recall next the basic definitions of ASP and splitting, to proceed with a formalization of epistemic splitting afterwards.

Background of ASP and splitting

Given a set of atoms At, a *regular literal*[1] is either an atom or constant, $a \in At \cup \{\top, \bot\}$, or its default negation, $\texttt{not}\ a$. A *rule* r is an implication of the form:

$$a_1 \vee \cdots \vee a_n \leftarrow L_1, \ldots, L_m \tag{9}$$

with $n \geq 0$ and $m \geq 0$, where each $a_i \in At$ is an atom and each L_j a regular literal. The left hand disjunction of (9) is called the rule *head* and abbreviated as $Head(r)$. When $n = 0$, it corresponds to $\bot$ and r is called a *constraint*. The right hand side of (9) is called the rule *body* and abbreviated as $Body(r)$. When $m = 0$, the body corresponds to $\top$ and r is called a *fact* (in this case, the body and the arrow symbol are usually omitted). A *program* Π is a set of rules. We write $Atoms(F)$ to represent the set of atoms occurring in any syntactic construct F (a literal, head, body, rule or program). A propositional interpretation I is a set of atoms. We assume that strong negation '$\sim a$' is just another atom in At and that the constraint:

$$\bot \leftarrow a, \sim a$$

is implicitly included in the program. We allow the use of variables, but understood as abbreviations of their possible ground instances. Given any syntactic construct F, we write $I \models F$ to stand for "I satisfies F" in classical propositional logic, where the commas correspond to conjunctions, '$\texttt{not}$' corresponds (under this interpretation) to classical negation and '$\leftarrow$' is just a reversed material implication. An interpretation I is a *(classical) model* of a program Π if it satisfies all its rules. The *reduct* of a program Π with respect to some propositional interpretation I, in symbols Π^I, is obtained by replacing in Π every negative literal $\texttt{not}\ a$ by $\top$ if $I \models \texttt{not}\ a$ or by $\bot$ otherwise. A propositional interpretation I is a *stable model* of a program Π iff it is a $\subseteq$-minimal model of Π^I. By $\mathrm{SM}[\Pi]$, we denote the set of all stable models of Π.

The following is a well-known property in ASP.

Property 1 (Supraclassicality)**.** *Any stable model of a program* Π *is also a classical model of* Π.

We recall next the splitting theorem for ASP, beginning with the following definition.

Definition 1 (Splitting set)**.** *A set of atoms* $U \subseteq \mathrm{At}$ *is a* splitting set *of a program* Π *if, for each rule* $r \in \Pi$*, one of the following conditions hold*

i) $Atoms(r) \subseteq U$,

ii) $Atoms(Head(r)) \cap U = \emptyset$

When this happens, we identify two disjoint subprograms, the bottom *and the* top*, respectively defined as*

$$b_U(\Pi) \stackrel{\text{def}}{=} \{\ r \in \Pi \mid Atoms(r) \subseteq U\ \}$$

$$t_U(\Pi) \stackrel{\text{def}}{=} \Pi \setminus b_U(\Pi) \qquad \square$$

As an example, consider program Π_1:

$$a \leftarrow \texttt{not}\ b \tag{10}$$

$$b \leftarrow \texttt{not}\ a \tag{11}$$

$$c \vee d \leftarrow \texttt{not}\ a \tag{12}$$

$$d \leftarrow a, \texttt{not}\ b \tag{13}$$

It is easy to see that $U = \{a, b\}$ is a splitting set that divides the program into two parts: the bottom $b_U(\Pi_1) = \{(10), (11)\}$ and the top $t_U(\Pi_1) = \{(12), (13)\}$.

The keypoint of splitting is computing stable models of $b_U(\Pi)$ alone and using each one, I, to simplify $t_U(\Pi)$ accordingly. Given a splitting set U for Π and an interpretation $I \subseteq U$, we define the program $e_U(\Pi, I)$ as a transformation of the top program, $t_U(\Pi)$, where we replace each atom $a \in U$ from the splitting set by: $\top$ if $a \in I$ or $\bot$ otherwise. A pair $\langle I_b, I_t \rangle$ is said to be a *solution* of Π with respect to U iff I_b is a stable model of $b_U(\Pi)$ and I_t is a stable model of $e_U(\Pi, I_b)$. For instance, for Π_1, the bottom has two answer sets $\{a\}$ and $\{b\}$, so we get the respective simplifications $e_U(\Pi_1, \{a\})$:

$$c \vee d \leftarrow \texttt{not}\ \top \qquad d \leftarrow \top, \texttt{not}\ \bot$$

and $e_U(\Pi_1, \{b\})$:

$$c \vee d \leftarrow \texttt{not}\ \bot \qquad d \leftarrow \bot, \texttt{not}\ \top$$

The former has stable model $\{d\}$ so $\langle \{a\}, \{d\} \rangle$ is one solution. The latter has stable models $\{c\}$ and $\{d\}$ that yield other two solutions $\langle \{b\}, \{c\} \rangle$ and $\langle \{b\}, \{d\} \rangle$.

Theorem 1 (From Lifschitz and Turner 1994)**.** *Let* U *be a splitting set of program* Π*. A propositional interpretation* $I \subseteq At$ *is a stable model of* Π *iff there is a solution* $\langle I_b, I_t \rangle$ *of* Π *w.r.t* U *such that* $I = I_b \cup I_t$*.* $\square$

Given the three solutions we obtained before, the splitting theorem guarantees that $\{a\} \cup \{d\}$, $\{b\} \cup \{c\}$ and $\{b\} \cup \{d\}$ are the three stable models of our example program Π_1.

One interesting observation is that any constraint r with $Atoms(r) \subseteq U$ is now included in the bottom $b_U(\Pi)$ but also satisfies condition ii) in Def. 1 (it has no head atoms at all) and could be moved to the top $t_U(\Pi)$ instead. Having this in mind, let us provide now a relaxed definition of bottom and top programs in the following way. We say that the pair $\langle \hat{b}_U(\Pi), \hat{t}_U(\Pi) \rangle$ is an *arbitrary splitting* of program Π with respect to splitting set U if: $\hat{b}_U(\Pi) \cap \hat{t}_U(\Pi) = \emptyset$, $\hat{b}_U(\Pi) \cup \hat{t}_U(\Pi) = \Pi$, all rules in B_U satisfy Def. 1.i) and all rules in T satisfy Def. 1.ii). With this definition, constraints on atoms in U can be arbitrarily placed in $\hat{b}_U(\Pi)$ or in $\hat{t}_U(\Pi)$.

Corollary 1. *Theorem 1 still holds if we define* $b_U(\Pi)$ *and* $t_U(\Pi)$ *to be any arbitrary splitting* $\langle \hat{b}_U(\Pi), \hat{t}_U(\Pi) \rangle$ *of program* Π *with respect to splitting set* U*.* $\square$

[1] For a simpler description of program transformations, we allow truth constants with their usual meaning.

Epistemic specifications

We extend now the syntax of ASP to the language of epistemic specifications. Given a set of atoms At, we call *subjective literal* to any expression of the form $\mathbf{K}\,l$, $\mathbf{M}\,l$, $\texttt{not}\ \mathbf{K}\,l$ or $\texttt{not}\ \mathbf{M}\,l$, for any regular literal l. We keep the same syntax for rules as in (9) excepting that body literals L_j can also be subjective literals now. Given rule r we define the sets $Body(r)^{reg}$ and $Body^{sub}(r)$ respectively containing the regular and the subjective literals in $Body(r)$. Rules or programs are *regular* if they do not contain subjective literals. We say that a rule is a *subjective constraint* if it is a constraint, $Head(r) = \bot$, and its body exclusively consists of subjective literals, that is $Body(r) = Body^{sub}(r)$.

We can define the concept of *model* of a program, in a similar way as we did for classical models in regular ASP. A *modal interpretation* $\mathcal{M} = \langle W, I\rangle$ is pair where I is a propositional interpretation and $W \subseteq 2^{At}$ is a non-empty set of propositional interpretations. A modal interpretation $\mathcal{M} = \langle W, I\rangle$ *satisfies* a literal L, written $\langle W, I\rangle \models L$, if

1. $\langle W, I\rangle \models \top$,
2. $\langle W, I\rangle \not\models \bot$,
3. $\langle W, I\rangle \models a$ if $a \in I$, for any atom $a \in At$,
4. $\langle W, I\rangle \models \mathbf{K}\,l$ if $\langle W, I'\rangle \models l$ for all $I' \in W$,
5. $\langle W, I\rangle \models \mathbf{M}\,l$ if $\langle W, I'\rangle \models l$ for some $I' \in W$, and
6. $\langle W, I\rangle \models \texttt{not}\ L$ if $\langle W, I\rangle \not\models L$.

Since for a subjective literal L, $\langle W, I\rangle \models L$ does not depend on I, we sometimes write $W \models L$. For a rule r of the form (9), we write $\langle W, I\rangle \models r$ iff either $\langle W, I\rangle \models a_i$ for some $1 \leq i \leq n$ or $\langle W, I\rangle \not\models L_j$ for some $1 \leq j \leq m$. We say that $\langle W, I\rangle$ is a *model* of a program Π, written $\langle W, I\rangle \models \Pi$, if it satisfies all its rules. Among the possible models of an epistemic logic program, all semantic approaches agree on selecting some preferred models called *world views*, each one being characterized by the W component. These world views satisfy a similar property to that of supraclassicality (Property 1) in non-epistemic ASP. In this case, however, rather than talking about classical models, we resort to modal logic S5, so all world views of a program are also S5 models of the program. This property can be formally stated as follows:

Property 2 (Supra-S5)**.** *A semantics satisfies* supra-S5 *when for every world view W of an epistemic program Π and for every $I \in W$, $\langle W, I\rangle \models \Pi$.* □

To the best of our knowledge, all existing semantics satisfy supra-S5. Another property that is shared by all semantics is that, when Π is a regular ASP program (it has no modal epistemic operators) then it has a unique world view containing all the stable models of Π. We will formalize this property in the following way.

Property 3 (Supra-ASP)**.** *A semantics satisfies* supra-ASP *if for any regular program Π either Π has a unique world view $W = \mathrm{SM}[\Pi] \neq \emptyset$ or $\mathrm{SM}[\Pi] = \emptyset$ and Π has no world view at all.* □

Originally, some semantics like (Gelfond, 1991) or (Truszczyński, 2011), allowed empty world views $W = \emptyset$ when the program has no stable models, rather than leaving the program without world views. Since this feature is not really essential, we exclusively refer to non-empty world views in this paper.

We define next a useful transformation extending the idea of reduct to epistemic specifications, and generalized for a given signature.

Definition 2 (Subjective reduct)**.** *The* subjective reduct *of a program Π with respect to a set of propositional interpretations W and a signature $U \subseteq \mathrm{At}$, also written Π^W_U, is obtained by replacing each subjective literal L with $Atoms(L) \subseteq U$ by: $\top$ if $W \models L$ or by $\bot$ otherwise. When $U = At$ we just write Π^W.* □

We use the same notation Π^W as for the standard reduct, but ambiguity is removed by the type of W (a set of interpretations now). This subjective reduct can be used to define (Gelfond, 1991) (G91) semantics in the following way.

Definition 3 (G91-world view)**.** *A non-empty set of interpretations W is a* G91-world view *of an epistemic program Π if $W = \mathrm{SM}[\Pi^W]$.* □

We will not provide the formal definitions of the rest of semantics compared in this paper, since none of them satisfies our goal property of epistemic splitting. In those cases, it will suffice with providing counterexamples and the reader can check their behaviour by resorting to the corresponding original definition.

Epistemic splitting

We proceed now to introduce our definition of the epistemic splitting property. To do so, we begin extending the idea of splitting set.

Definition 4 (Epistemic splitting set)**.** *A set of atoms $U \subseteq \mathrm{At}$ is said to be an* epistemic splitting set *of a program Π if for any rule r in Π one of the following conditions hold*

i) $Atoms(r) \subseteq U$,

ii) $(Atoms(Body^{reg}(r) \cup Head(r))) \cap U = \emptyset$

As before, we define an arbitrary splitting *of Π as a pair $\langle B_U(\Pi), T_U(\Pi)\rangle$ satisfying $B_U(\Pi) \cap T_U(\Pi) = \emptyset$, $B_U(\Pi) \cup T_U(\Pi) = \Pi$, all rules in $B_U(\Pi)$ satisfy i) and all rules in $T_U(\Pi)$ satisfy ii).* □

With respect to Definition 1, we have replaced the condition for the top program, $Atoms(Head(r)) \cap U = \emptyset$, by the new condition ii), which in other words means that the top program may only refer to atoms U in the bottom through epistemic operators. Note that this introduces a new kind of "dependence," so that, as happens with head atoms, regular literals in the body also depend on atoms in subjective literals. For instance, if $U = \{p, q\}$, the program

$$p \vee q \qquad\qquad s \leftarrow p, \mathbf{K}\,q$$

would not be splittable due to the second rule, since $s \notin U$ and we would also need the regular literal $p \notin U$. The reason for this restriction is to avoid imposing (to a potential semantics) a fixed way of evaluating p with respect to the world view $[\{p\}, \{q\}]$ for the bottom.

Another observation is that we have kept the definition of $B_U(\Pi)$ and $T_U(\Pi)$ non-deterministic, in the sense that some rules can be arbitrarily included in one set or the other. In our case, these rules correspond to subjective constraints on atoms in U, since these are the only cases that may satisfy conditions i) and ii) simultaneously.

If we retake our example program $\Pi_2 = \{(1) - (5)\}$, we can see that $U = \{high(mike), fair(mike), eligible(mike), minority(mike)\}$ is an epistemic splitting set that divides the program into the bottom $B_U(\Pi_2) = \{(1) - (4)\}$ and the top $T_U(\Pi_2) = \{(5)\}$. As in regular splitting, the idea is computing first the world views of the bottom program $B_U(\Pi)$ and for each one, W, simplifying the corresponding subjective literals in the top program. Given an epistemic splitting set U for a program Π and set of interpretations W, we define $E_U(\Pi, W) \stackrel{\text{def}}{=} T_U(\Pi)^W_U$, that is, we make the subjective reduct of the top with respect to W and signature U. A pair $\langle W_b, W_t \rangle$ is said to be a *solution* of Π with respect to an epistemic splitting set U if W_b is a world view of $B_U(\Pi)$ and W_t is a world view of $E_U(\Pi, W_b)$. Notice that this definition is semantic-dependent in the sense that each alternative semantics for epistemic specifications will define its own solutions for a given U and Π, since it defines the selected world views for a program in a different way. Back to our example, notice that $B_U(\Pi_2)$ is a regular program without epistemic operators. Thus, any semantics satisfying supra-ASP will provide $W_b = [\{fair(mike)\}, \{high(mike), eligible(mike)\}]$ as unique world view for the bottom. The corresponding simplification of the top would be $E_U(\Pi_s, W_b)$ containing (after grounding) the single rule:

$$interview(mike) \leftarrow \texttt{not}\ \bot, \texttt{not}\ \bot$$

Again, this program is regular and its unique world view would be $W_t = [\{interview(mike)\}]$. Now, in the general case, to reconstruct the world views for the global program we define the operation:

$$W_b \sqcup W_t \quad = \quad \{\ I_b \cup I_t \mid I_b \in W_b \text{ and } I_t \in W_t\ \}$$

(remember that both the bottom and the top may produce multiple world views, depending on the program and the semantics we choose). In our example, $W_b \sqcup W_t$ would exactly contain the two stable models $(6), (7)$ we saw in the introduction.

Property 4 (Epistemic splitting)**.** *A semantics satisfies* epistemic splitting *if for any epistemic splitting set U of any program Π: W is a world view of Π iff there is a solution $\langle W_b, W_t \rangle$ of Π with respect to U such that $W = W_b \sqcup W_t$.* □

In the example, this means that the world view we obtained in two steps is indeed the unique world view of the whole program, under any semantics satisfying epistemic splitting. Uniqueness of world view was obtained in this case because both the bottom program $B_U(\Pi_2)$ and the top, after simplification, $E_U(\Pi_2, W_b)$ were regular programs and we assumed supra-ASP. In fact, as we see next, we can still get a unique world view (at most) when there are no cyclic dependences among subjective literals. This mimics the well-known result for *stratified negation* in logic programming (van Gelder, 1988; Apt, Blair, and Walker, 1988). Let us define a modal dependence relation among atoms in a program Π so that $dep(a, b)$ is true iff there is a rule $r \in \Pi$ such that $a \in Atoms(Head(r) \cup Body^{reg}(r))$ and $b \in Atoms(Body^{sub}(r))$.

Definition 5. *We say that an epistemic program Π is* modally stratified *if we can assign an integer mapping $\lambda : \text{At} \to \mathbb{N}$ to each atom such that $\lambda(a) > \lambda(b)$ for any pair of atoms a, b satisfying $dep(a, b)$.* □

Take, for instance, the extended program $\Pi_3 = \{(1) - (5), (8)\}$. We can assign atoms $high(mike)$, $fair(mike)$, $minority(mike)$ and $eligible(mike)$ layer 0. Then $interview(mike)$ could be assigned layer 1 and, finally, $appointment(mike)$ can be located at layer 2. So, Π_3 is modally stratified.

Theorem 2. *Let Π be a finite, modally stratified program. Then, any semantics satisfying supra-ASP and epistemic splitting assigns, at most, a unique world view to Π.* □

The proof of the theorem just relies on multiple applications of splitting to each layer backwards and the fact that each simplification $E_U(\Pi, W_b)$ will be a regular program. This is very easy to see in the extended example Π_3. We can split the program using as U all atoms but $appointment(mike)$ to get a bottom Π_2 and a top $\{(8)\}$. Program Π_2 can be splitted in its turn as we saw before, producing the unique world view $\{(6), (7)\}$. Then $E_U(\Pi_3, \{(6), (7)\})$ contains the single rule $appointment(mike) \leftarrow \top$ that is a regular program whose unique world view is $[\{appointment(mike)\}]$ and, finally, the combination of both world views yields again a unique world view $[(6) \cup \{appointment(mike)\}, (7) \cup \{appointment(mike)\}]$.

Another consequence of epistemic splitting is that subjective constraints will have a monotonic behaviour. Note first that, for a subjective constraint r, we can abbreviate $\langle W, I \rangle \models r$ as $W \models r$ since the I component is irrelevant. Additionally, $W \models r$ means that $Body(r) = Body^{sub}(r)$ is falsified, since $Head(r) = \bot$.

Property 5 (subjective constraint monotonicity)**.** *A semantics satisfies* subjective constraint monotonicity *if, for any epistemic program Π and any subjective constraint r, W is a world view of $\Pi \cup \{r\}$ iff both W is a world view of Π and $W \models r$.* □

Theorem 3. *Epistemic splitting implies subjective constraint monotonicity.* □

Proof. Suppose we use a semantics satisfying epistemic

splitting. For any program Π and any epistemic constraint r, we can always take the whole set of atoms $U = \mathit{Atoms}(\Pi \cup \{r\})$ as epistemic splitting set for $\Pi' = \Pi \cup \{r\}$ and take $B_U(\Pi') = \Pi$ and $T_U(\Pi') = \{r\}$. For any world view W of $B_U(\Pi')$ two things may happen. A first possibility is $W \models r$, and so the body of r has some false subjective literal in W, so $E_U(\Pi', W)$ would be equivalent to $\bot \leftarrow \bot$. Then, the unique world view for the top would be $W_t = [\emptyset]$ and $W \sqcup W_t = W$. A second case is $W \not\models r$, so all literals in the body are satisfied and $E_U(\Pi', W)$ would be equivalent to $\bot \leftarrow \top$ which has no world views. To sum up, we get exactly those world views W of Π that satisfy r. □

To conclude the exploration of consequences of epistemic splitting, let us consider a possible application to conformant planning. To this aim, consider the following simple example.

Example 2. *To turn on the light in a room, we can toggle one of two lamps l_1 or l_2. In the initial state, lamp l_1 is plugged but we ignore the state of l_2. Our goal is finding a plan that guarantees we get light in the room in one step.*

A possible logic program that encodes this scenario for a single transition[2] could be Π_4:

$$\mathit{plugged}(l_1)$$
$$\mathit{plugged}(l_2) \vee {\sim}\mathit{plugged}(l_2)$$

$$\begin{array}{rcl} \mathit{light} & \leftarrow & \mathit{toggle}(L), \mathit{plugged}(L) \\ \bot & \leftarrow & \mathit{toggle}(l_1), \mathit{toggle}(l_2) \end{array}$$

for $L \in \{l_1, l_2\}$. As we can see, $\mathit{toggle}(l_1)$ would constitute a conformant plan, since we obtain *light* regardless of the initial state, while this does not happen with plan $\mathit{toggle}(l_2)$. In order to check whether a given sequence of actions $A_0, \ldots, A_n$ is a valid conformant plan one would expect that, if we added those facts to the program, a subjective constraint should be sufficient to check that the goal holds in all the possible outcomes. In our example, we would just use:

$$\bot \leftarrow \texttt{not}\ \mathbf{K}\,\mathit{light} \tag{14}$$

and check that the program $\Pi_4 \cup \{\mathit{toggle}(L)\} \cup \{(14)\}$ has some world view, varying $L \in \{l_1, l_2\}$. Subjective constraint monotonicity guarantees that the addition of this "straighforward" formalisation has the expected meaning.

This method would only allow testing if the sequence of actions constitutes a conformant plan, but does not allow generating those actions. A desirable feature would be the possibility of applying the well-known ASP methodology of separating the program into three sections: generate, define and test. In our case, the "define" and the "test" sections would respectively be Π_4 and (14), but we still miss a "generate" part, capable of considering different alternative conformant plans. The problem in this case is that we cannot use a simple choice:

$$\mathit{toggle}(L) \vee {\sim}\mathit{toggle}(L)$$

because this would allow a same action to be executed in some of the stable models and not executed in others, all inside a *same* world view. Let us assume that our epistemic semantics has some way to non-deterministically generate a world view in which either $\mathbf{K}\,a$ or $\mathbf{K}\,\texttt{not}\ a$ holds using a given set of rules[3] $\mathit{Choice}(a)$. Then, take the program Π_5 consisting of rules

$$\mathit{Choice}(\mathit{toggle}(L)) \tag{15}$$

with $L \in \{l_1, l_2\}$ plus Π_4 and (14). If our semantics satisfies epistemic splitting, it is safe to obtain the world views in three steps: generate first the alternative world views for $\mathit{toggle}(l_1)$ and $\mathit{toggle}(l_2)$ using (15), apply Π_4 and rule out those world views not satisfying the goal *light* in all situations using (14). To fulfill the preconditions for applying splitting, we would actually need to replace regular literal $\mathit{toggle}(L)$ by $\mathbf{K}\,\mathit{toggle}(L)$ in all the bodies of Π_4, but this is safe in the current context. Now, we take the bottom program to obtain 4 possible world views $W_0 = [\{\mathit{toggle}(l_1)\}]$, $W_1 = [\{\mathit{toggle}(l_2)\}]$, $W_2 = [\{\mathit{toggle}(l_1), \mathit{toggle}(l_2)\}]$ and $W_3 = [\emptyset]$. When we combine them with the top Π_4 we obtain W'_0 consisting of two stable models:

$$\{\mathit{toggle}(l_1), \mathit{plugged}(l_2), \mathit{light}, \ldots\}$$
$$\{\mathit{toggle}(l_1), {\sim}\mathit{plugged}(l_2), \mathit{light}, \ldots\}$$

and W'_1 consisting of other two stable models:

$$\{\mathit{toggle}(l_2), \mathit{plugged}(l_2), \mathit{light}, \ldots\}$$
$$\{\mathit{toggle}(l_2), {\sim}\mathit{plugged}(l_2), \ldots\}$$

where the latter does not contain *light*. Finally, constraint (14) would rule out W'_1.

To sum up, epistemic splitting provides a natural way of formulating conformant planning problems by a separation into three sections: a generation part, the usual encoding of the actions scenario and a test part consisting of a subjective constraint to guarantee that the goal is always reached.

Splitting in some existing semantics

In this section we study the property of epistemic splitting for the approaches mentioned in the introduction. We will begin by proving that G91 actually satisfies this property. To this aim, we start with some definitions and auxiliary results. Given a set of propositional interpretations $W \subseteq 2^{At}$ and a set of atoms U, by $W_{|U} \stackrel{\text{def}}{=} \{\ I \cap U \mid I \in U\ \}$, we denote the restriction of W to U. Given a set of atoms U, by $\overline{U}$, we denote its complement $At \setminus U$.

[2]For simplicity, we omit time arguments or inertia, as they are not essential for the discussion.

[3]For instance, in Gelfond (1991), this could be just the rule $a \leftarrow \texttt{not}\ \mathbf{K}\,\texttt{not}\ a$. Other semantics may have alternative ways of expressing this intended behaviour.

Observation 1. *Let W be a set of propositional interpretations and $U \subseteq \mathrm{At}$ be a set of atoms. Then, for any subjective literal L with $Atoms(L) = \{a\}$:*

i) if $a \in U$, then $W \models L$ iff $W_{|U} \models L$,

ii) if $a \notin U$, then $W \models L$ iff $W_{|\overline{U}} \models L$,

Proposition 1. *Let Π be a program that accepts an epistemic splitting set $U \subseteq At$ and let W be a set of propositional interpretations. Let $W_b = W_{|U}$ and $W_t = W_{|\overline{U}}$. Then, we get*

i) $B_U(\Pi)^W = B_U(\Pi)^{W_b}$,

ii) $T_U(\Pi)^W = E_U(\Pi, W_b)^{W_t}$, and

iii) $\Pi^W = B_U(\Pi)^{W_b} \cup E_U(\Pi, W_b)^{W_t}$.

Proof. First, since every rule $r \in B_U(\Pi)$ satisfies $Atoms(Body^{sub}(r)) \subseteq U$, from Observation 1, it follows that $B_U(\Pi)^W = B_U(\Pi)^{W_b}$. Furthemore, for any program Γ, it is easy to check that $\Gamma^W = (\Gamma_U^{W_b})^{W_t}$, that is, applying the reduct w.r.t W is the same than applying it w.r.t. to its projection in U and afterwards to the remaining part. Thus, we get $T_U(\Pi)^W = (T_U(\Pi)_U^{W_b})^{W_t} = E_U(\Pi, W_b)^{W_t}$. Finally, we have that $\Pi^W = (B_U(\Pi) \cup T_U(\Pi))^W = B_U(\Pi)^W \cup T_U(\Pi)^W$ and, thus, the result holds. □

Theorem 4 (Main theorem)**.** *Semantics G91 satisfies epistemic splitting.*

Proof. Let W be some set of propositional interpretations and let $W_b = W_{|U}$ and $W_t = W_{|\overline{U}}$. By definition, W is a world view of Π if and only if $W = \mathrm{SM}[\Pi^W]$. Furthermore, since U is a modal splitting set of Π, it is easy to check that U is also a regular splitting set of the regular programa Π^W. Hence, from Corollary 1, we get that W is a world view of Π iff $W = \mathrm{SM}[\Pi^W] =$

$$\{\ I_b \cup I_t \ \mid\ I_b \in \mathrm{SM}[\hat{b}_U(\Pi^W)] \text{ and } I_t \in \mathrm{SM}[\hat{e}_U(\Pi^W, I_b)]\ \}$$

for some arbitrary splitting $\langle \hat{b}_U(\Pi^W), \hat{t}_U(\Pi^W)\rangle$. Note that all rules belonging to $B_U(\Pi)$ have all atoms from U. Hence, we take $\hat{b}_U(\Pi^W) \stackrel{\text{def}}{=} B_U(\Pi)^W = B_U(\Pi)^{W_b}$ (Proposition 1). Similarly, we also take $\hat{t}_U(\Pi^W) \stackrel{\text{def}}{=} T_U(\Pi)^W = E_U(\Pi, W_b)^{W_t}$. Then, we get

$$\hat{e}_U(\Pi^W, I_b) \ = \ \hat{e}_U(\hat{t}_U(\Pi^W), I_b) \ = \ \hat{e}_U(E_U(\Pi, W_b)^{W_t}, I_b)$$

Notice also that no atom occurring in $E_U(\Pi, W_b)^{W_t}$ belongs to U, which implies that $\hat{e}_U(E_U(\Pi, W_b)^{W_t}, I_b) = E_U(\Pi, W_b)^{W_t}$. Replacing above, we have that W is a world view of Π iff W is equal to

$$\{I_b \cup I_t \mid I_b \in \mathrm{SM}[B_U(\Pi)^{W_b}], I_t \in \mathrm{SM}[E_U(\Pi, W_b)^{W_t}]\}$$

iff

$$W = \{I_b \cup I_t \mid I_b \in W'_b \text{ and } I_t \in W'_t\}$$

with $W'_b = \mathrm{SM}[B_U(\Pi)^{W_b}]$ and $W'_t = \mathrm{SM}[E_U(\Pi, W_b)^{W_t}]$ iff $W = W'_b \sqcup W'_t$. Hence, it only remains to be shown that both $W_b = W'_b$ and $W_t = W'_t$ hold. Note that $I \in W_b = W_{|U}$ iff $I = I' \cap U$ for some $I' \in W$ iff $I = (I_b \cup I_t) \cap U$ for some $I_b \in W'_b$ and $I_t \in W'_t$ iff $I = (I_b \cap U) \cup (I_t \cap U)$ for some $I_b \in W'_b$ and $I_t \in W'_t$ iff $I = I_b$ for some $I_b \in W'_b$. The fact $W_t = W'_t$ follows in an analogous way. □

A similar proof can be developed to show that (Truszczyński, 2011), that generalises[4] (Gelfond, 1991) from subjective literals to subjective formulas, also satisfies epistemic splitting.

To illustrate the behaviour of other semantics with respect to splitting, we will use several examples. Let us take the program Π_6 consisting of $\{(10), (11)\}$ and the rule:

$$c \vee d \ \leftarrow \ \texttt{not}\ \mathbf{K}\, a \tag{16}$$

The set $U = \{a, b\}$ splits the program into the bottom, (10)-(11) and the top (16). The bottom has a unique world view $W_b = [\{a\}, \{b\}]$ so $\mathbf{K}\, a$ does not hold and the top is simplified as $E_U(\Pi_6, W_b)$ containing the unique rule:

$$c \vee d \leftarrow \texttt{not}\ \bot \tag{17}$$

This program has a unique world view $W_t = [\{c\}, \{d\}]$ that, combined with W_b yields $[\{a, c\}, \{b, c\}, \{a, d\}, \{b, d\}]$ as the unique solution for Π_6, for any semantics satisfying epistemic splitting (and so, also for G91). Let us elaborate the example a little bit further. Suppose we add now the constraint:

$$\bot \leftarrow c \tag{18}$$

The top must also include this rule and has now a unique stable model $W_t = [\{d\}]$, so the world view for the complete program would be $[\{a, d\}, \{b, d\}]$. Finally, let us forbid the inclusion of atom d too:

$$\bot \leftarrow d \tag{19}$$

so we consider $\Pi_7 = \{(10), (11), (16), (18), (19)\}$. This last constraint leaves the simplified top program $E_U(\Pi_6, W_b) = \{(17), (18), (19)\}$ without stable models, so epistemic splitting would yield that program Π_7 has no world view at all. This is the result we obtain, indeed, in (Gelfond, 1991, 2011)[5] and in (Truszczyński, 2011). Surprisingly, recent approaches like (Kahl et al., 2015; Fariñas del Cerro, Herzig, and Su, 2015; Shen and Eiter, 2017; Son et al., 2017) yield world view $[\{a\}]$, violating the epistemic splitting property. For instance, in the case of (Kahl et al., 2015), the reduct of Π_7 with respect to $[\{a\}]$ is the program

$$\begin{array}{rcl} a & \leftarrow & \texttt{not}\ b \\ b & \leftarrow & \texttt{not}\ a \\ c \vee d & \leftarrow & \texttt{not}\ a \\ \bot & \leftarrow & c \\ \bot & \leftarrow & d \end{array}$$

[4] In fact, Truszczyński (2011) defines several semantics but, among them, we refer here to the epistemic stable model semantics.

[5] These two semantics actually produce empty world views, but as we said before, we disregard them, as they just point out that the program has no solution.

which has a unique stable model $\{a\}$.

As a second example, take the program Π_8 consisting of the same bottom program $\{(10), (11)\}$ and the rule:

$$c \leftarrow \mathbf{K}\, a \tag{20}$$

As expected, all approaches agree that Π_8 has a unique world view $W_b = [\{a\}, \{b\}]$ because $\mathbf{K}\, a$ is not satisfied and rule (20) is not applicable. Under epistemic splitting, we get that $E_U(\Pi_8, W_b)$ is the rule:

$$c \leftarrow \bot \tag{21}$$

whose unique world view is $[\emptyset]$, so that $W_b \sqcup [\emptyset] = W_b$. But let us further elaborate the example taking Π'_8 containing Π_8 plus:

$$\bot \leftarrow \texttt{not}\ c \tag{22}$$

Under epistemic splitting, the new top $E_U(\Pi'_8, W_b)$ contains now (22) and (21) which have no stable models. As a result, no world view can be combined with W_b and we obtain that Π'_8 has no world views at all. This is the result we obtain under (Gelfond, 1991; Truszczyński, 2011), which agree that the program is inconsistent. However, Gelfond (2011) joins (Kahl et al., 2015; Fariñas del Cerro, Herzig, and Su, 2015; Shen and Eiter, 2017; Son et al., 2017) in the group of approaches that provide the world view $[\{a, c\}]$. That is, in all these approaches, adding a constraint intended to remove all world views that do not satisfy c, may surprisingly lead to justify c. Note that, according to (Gelfond, 1991; Truszczyński, 2011), the reduct of Π'_8 with respect to $[\{a, c\}]$ is

$$a \leftarrow \texttt{not}\ b \tag{23}$$
$$b \leftarrow \texttt{not}\ a \tag{24}$$
$$c \leftarrow \top \tag{25}$$
$$\bot \leftarrow \texttt{not}\ c \tag{26}$$

which has two stable models, $\{a, c\}$ and $\{b, c\}$, so $[\{a, c\}]$ is not a world view. In contrast, the reduct with respect to (Gelfond, 2011) and (Kahl et al., 2015) is

$$a \leftarrow \texttt{not}\ b \tag{27}$$
$$b \leftarrow \texttt{not}\ a \tag{28}$$
$$c \leftarrow a \tag{29}$$
$$\bot \leftarrow \texttt{not}\ c \tag{30}$$

which has a unique stable model $\{a, c\}$, so $[\{a, c\}]$ is a world view.

Conclusions

We have introduced a formal property for semantics of epistemic specifications. This property that we call *epistemic splitting* has a strong resemblance to the splitting theorem well-known for regular ASP programs. Epistemic splitting can be applied when we can divide an epistemic logic program into a bottom part for a subset U of atoms and a top part, that only refers to atoms in U through subjective literals (those using modal epistemic operators). When this happens, the property of splitting states that we should be able to compute the world views of the program in two steps: first, computing the world views of the bottom and, second, using each bottom world view W to replace subjective literals for atoms in U in the top by their truth value with respect to W.

We have studied several consequences of epistemic splitting: for instance, if the program is stratified with respect to subjective literals then it will have a unique world view, at most. Another consequence is that constraints only consisting of subjective literals will have a monotonic behaviour, ruling out world views that satisfy the constraint body.[6] We have also explored how epistemic splitting may facilitate the simple application of the generate-define-test methodology, well-known in ASP, to the formalisation of conformant planning.

Our study of the main semantics in the literature has shown that only the original semantics (Gelfond, 1991) (G91), and its generalisation (Truszczyński, 2011), satisfy epistemic splitting while the rest of approaches we considered no, as we showed with counterexamples. As said in the introduction, a different kind of epistemic splitting had also been proved for G91 in (Watson, 2000), reinforcing the idea that this semantics can be interpreted in a modular way. Notice that the sets of programs that can be split under these two definitions is incomparable. We do not mean with this, however, that G91 is always intuitive. As it is well-known, G91 suffers from self-supportedness: for instance, the program consisting of the single rule $p \leftarrow \mathbf{K}\, p$ yields two world views $[\emptyset]$ and $[\{p\}]$ but the latter justifies p by the mere assumption of $\mathbf{K}\, p$ without further evidence, something that seems counterintuitive. What we claim instead is that G91 has a reasonable behaviour when subjective literals are stratified. Unfortunately, later attempts to solve self-supportedness on cyclic epistemic specifications have somehow spoiled that feature.

References

Apt, K. R.; Blair, H. A.; and Walker, A. 1988. Towards a theory of declarative knowledge. In *Foundations of Deductive Databases and Logic Programming.* Morgan Kaufmann. 89–148.

Fariñas del Cerro, L.; Herzig, A.; and Su, E. I. 2015. Epistemic equilibrium logic. In *Proc. of the Intl. Joint Conference on Artificial Intelligence (IJCAI'15)*, 2964–2970. AAAI Press.

Gelfond, M., and Lifschitz, V. 1988. The stable model semantics for logic programming. In *Proc. of the 5th Intl. Conference on Logic Programming (ICLP'88)*, 1070–1080.

Gelfond, M. 1991. Strong introspection. In Dean, T. L., and McKeown, K., eds., *Proceedings of the AAAI*

[6] The lack of monotonicity suffered by epistemic constraints in some semantics has been recently discussed by Leclerc and Kahl (2018a).

Conference, volume 1, 386–391. AAAI Press/The MIT Press.

Gelfond, M. 2011. New semantics for epistemic specifications. In *LPNMR*, volume 6645 of *Lecture Notes in Computer Science*, 260–265. Springer.

Kahl, P.; Watson, R.; Balai, E.; Gelfond, M.; and Zhang, Y. 2015. The language of epistemic specifications (refined) including a prototype solver. *Journal of Logic and Computation*.

Kahl, P. T. 2014. *Refining the semantics for epistemic logic programming*. Ph.D. Dissertation.

Leclerc, A. P., and Kahl, P. T. 2018a. Epistemic logic programs with world view constraints. In *Technical communication, 34th International Conference on Logic Programming (ICLP'2018)*.

Leclerc, A. P., and Kahl, P. T. 2018b. A survey of advances in epistemic logic program solvers. In *Proc. of the 11th Intl. Workshop on Answer Set Programming and other Computer Paradigms (ASPOCP'18)*.

Lifschitz, V., and Turner, H. 1994. Splitting a logic program. In *Proc. of the Intl. Conference on Logic Programming (ICLP'94)*, 23–37. MIT Press.

Shen, Y., and Eiter, T. 2017. Evaluating epistemic negation in answer set programming (extended abstract). In *Proc. of the Intl. Joint Conference on Artificial Intelligence (IJCAI'17)*, 5060–5064.

Son, T. C.; Le, T.; Kahl, P. T.; and Leclerc, A. P. 2017. On computing world views of epistemic logic programs. In *Proc. of the Intl. Joint Conference on Artificial Intelligence (IJCAI'15)*, 1269–1275. ijcai.org.

Truszczyński, M. 2011. Revisiting epistemic specifications. In *Logic Programming, Knowledge Representation, and Nonmonotonic Reasoning*, volume 6565 of *Lecture Notes in Computer Science*, 315–333. Springer.

van Gelder, A. 1988. Negation as failure using tight derivations for general logic programs. In *Foundations of Deductive Databases and Logic Programming*. Morgan Kaufmann. 149–176.

Watson, R. 2000. A splitting set theorem for epistemic specifications. *CoRR: Proceedings of the 8th International Workshop on Non-Monotonic Reasoning, NMR 2000* cs.AI/0003038.