Basic influence diagrams and the liberal stable semantics

Paul-Amaury MATT and Francesca TONI
Imperial College London, Department of Computing

Abstract. This paper is concerned with the general problem of constructing decision tables and more specifically, with the identification of all possible outcomes of decisions. We introduce and propose *basic influence diagrams* as a simple way of describing problems of decision making under strict uncertainty. We then establish a correspondence between basic influence diagrams and symmetric generalised assumption-based argumentation frameworks and adopt an argumentation-based approach to identify the possible outcomes. We show that the intended solutions are best characterised using a new semantics that we call *liberal stability*. We finally present a number of theoretical results concerning the relationships between liberal stability and existing semantics for argumentation.

Keywords. Abstract and assumption-based argumentation, decision tables and decision making.

Introduction

In decision theory [French, 1987], a decision making problem is modelled in terms of a decision table representing the outcomes of decisions under different scenarios, a probability distribution representing the likelihood of occurrence of the scenarios and a utility function measuring the subjective value of each possible outcome. In general, the construction of a decision table is not straightforward, for several reasons. Firstly, one must choose a way of representing the decisions, scenarios and possible outcomes. Secondly, finding an exhaustive set of disjoint scenarios and being able to describe the outcome of each decision under each one of them is a long process which requires expert knowledge in the practical domain of application. Thirdly, identifying the possible outcomes of decisions implicitly involves reasoning in the presence of conflicts that capture what is materially impossible or epistemically inconceivable. This paper presents a paradigm based on logical argumentation for finding all possible outcomes of decisions and constructing decision tables.

This approach is justified insofar as logical argumentation allows to represent decisions, scenarios and outcomes using literals, to capture their logical dependencies, to reason about them and resolve conflicts in a rational way. Concretely, we adopt *assumption-based argumentation* [Bondarenko *et al.*, 1997; Dung *et al.*, 2006; Toni, 2007] for identifying the possible outcomes of decisions, given *basic influence diagrams*, a simple tool for describing problems of decision making under strict uncertainty. Diagrams which are very similar in essence have already been deployed to feed into (a different form of) argumentation in [Morge and Mancarella, 2007].

The paper is organised as follows. Section 1 introduces basic influence diagrams. Section 2 recalls background definitions for assumption-based argumentation and introduces a generalisation needed for the purposes of this paper. Section 3 shows how to transform basic influence diagrams into argumentation frameworks. Then section 4 presents and justifies the conditions under which the solutions of basic influence diagrams are "rational". Section 5 shows that existing semantics of argumentation for conflict resolution mismatch with these conditions and proposes the *liberal stable semantics* as an alternative. The properties of liberal stability and its relationships with other semantics are then studied in detail. Section 6 summarises and concludes the paper.

1. Basic influence diagrams

The problem we study can be summarised by the following question:

What are all the possible outcomes of our decisions in a given decision domain ?

In order to construct a decision table, it is necessary to identify all possible outcomes and not just simply the best or most likely ones. We introduce *basic influence diagrams* to model the decision maker's goals, decisions, uncertainties and beliefs, their causal dependencies and possible conflicts. These diagrams are very similar in essence to Bayesian (or belief) networks [Pearl, 1986] and influence diagrams [Howard and Matheson, 1981] but are simpler, non-numerical data structures. Such diagrams are widely used for knowledge representation in artificial intelligence and recently, simpler qualitative forms of diagrams have started to be used to structure argumentation-based systems for decision support [Morge and Mancarella, 2007]. Formally, a *basic influence diagram* is composed of two parts: a diagram and a set of dependency rules. The diagram is an annotated finite directed acyclic graph whose nodes are literals of the form p or $\neg p$ where p is a proposition in some given language $\mathcal{L}$. Every literal belongs to one of the following categories:

- *Goals* are the nodes that have no outgoing arcs. Goals represent what the decision maker ultimately desires to achieve (positive goals) or avoid (negative goals). Positive/negative goals are graphically distinguished with a +/- symbol.
- *Atomic decisions* are an arbitrary strict subset of the nodes that are not goals and that have no incoming arcs. Atomic decisions are graphically distinguished from the other nodes by a squared box.
- *Beliefs* are the nodes which are neither goals nor decisions. Beliefs do not have any particular distinguishing feature. The beliefs that have no incoming arcs are called *fundamental beliefs*. The fundamental beliefs that are known to be true are called *facts* and are underlined. The other fundamental beliefs are called *unobservable fundamental beliefs* and are annotated with a ? symbol.

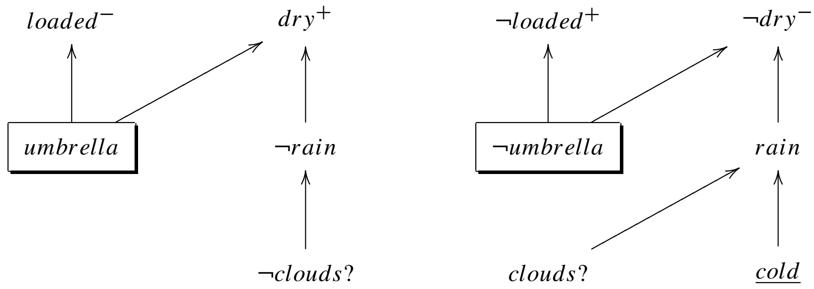
An arc from node p to node q means that the truth of the literal q logically depends on the truth of the literal p. Logical dependencies are expressed by *dependency rules*, which may be of the form

- "if p_1 and ... and p_k then q", or
- "q" as fact or observation

where p_1,...,p_k and q are literals. The arcs of the influence diagram and the dependency rules must conform in the following way. For each rule of the form "if p_1 and ...

and p_k then q" there must exist an arc from p_i to q for every $i \in \{1, ..., k\}$. For each dependency rule "q", q must appear as a fact in the diagram. Conversely, for each arc from p to q in the diagram, there must be at least one dependency rule "if ... p ... then q", and for each fact q there must exist a dependency rule "q".

Let us now see how to concretely represent a decision making problem using a basic influence diagram. The basic influence diagram[1] shown in figure 1 models a simple decision problem borrowed and adapted from [Amgoud and Prade, 2004], in which one has to decide whether to take an umbrella or not when going out for a walk. The decision maker has some elementary knowledge about the weather. He knows that the weather is cold, and he knows that if it is cold and there are clouds, then it must be raining. Taking an umbrella guarantees him to be able to stay dray but has the disadvantage of making him loaded.



if $umbrella$ then $loaded$
if $umbrella$ then dry
if $\neg rain$ then dry
if $\neg umbrella$ then $\neg loaded$
if $\neg umbrella$ and $rain$ then $\neg dry$
if $\neg clouds$ then $\neg rain$
if $clouds$ and $cold$ then $rain$
$cold$

Figure 1. Basic influence diagram corresponding to the umbrella example.

Dependency rules are purely deterministic, unlike probabilistic dependencies in Bayesian networks. In basic influence diagrams, uncertainties are not quantified, as probabilities play no role in the identification of the possible outcomes. Similarly, basic influence diagrams do not have a utility/value node as in standard influence diagrams. Decision tables should display all possible outcomes and not just those with maximal utility. Moreover, a utility function may only be defined once the set of outcomes has been identified. Although diagrams are not essential for representing knowledge in decision making under strict uncertainty they are nevertheless an excellent graphical aid to the specification of dependency rules.

Our final remark on basic influence diagrams concerns the use of negation $\neg$. First, when a literal p is a node of the diagram, its negation $\neg p$ is not required to be in the diagram. In the example, $\neg cold$ is not a literal of interest. Second, whenever a literal in

[1] In this example, the literal/node dry is true when either of its predecessors $umbrella$ or $\neg rain$ is true. On the opposite, the literal/node $\neg dry$ is true when both of its predecessors $\neg umbrella$ and $rain$ are true. This is not implied by the diagram and justifies the use of dependency rules.

the diagram admits a negation, one must make sure that practically, in the decision domain, either p holds or $\neg p$ holds. Third, cases where both p and $\neg p$ hold are considered as flawed or invalid opinions about the situation. Finally, cases where neither p nor $\neg p$ hold are considered as incomplete opinions about the situation. Therefore, the user of basic influence diagrams shall be aware that strong logical dependencies exist between nodes of the form p and $\neg p$.

2. Generalised assumption-based argumentation

This section provides background on assumption-based argumentation [Bondarenko *et al.*, 1997; Dung *et al.*, 2006; Toni, 2007; Dung *et al.*, 2007] and introduces a simple generalisation of this form of argumentation. An *assumption-based argumentation framework* is a tuple $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$ where

- $(\mathcal{L}, \mathcal{R})$ is a *deductive system*, consisting of a language $\mathcal{L}$ and a set $\mathcal{R}$ of inference rules,
- $\mathcal{A} \subseteq \mathcal{L}$, referred to as the set of *assumptions*,
- $\mathcal{C}$:$\mathcal{A} \mapsto \mathcal{L}$ is a total mapping associating a *contrary* sentence to each assumption.

The elements of $\mathcal{L}$ are called sentences. The inference rules in $\mathcal{R}$ have the syntax $\frac{p_1,\cdots p_n}{q}$ (for $n \geq 0$) where $p_i \in \mathcal{L}$. We will refer to q and $p_1, \ldots p_n$ as the *conclusion* and the *premises* of the rule, respectively. We restrict our attention to *flat* frameworks, such that if $q \in \mathcal{A}$, then there exists no inference rule of the form $\frac{p_1,\ldots,p_n}{q} \in \mathcal{R}$, for any $n \geq 0$.

A *generalised assumption-based argumentation framework* is an assumption-based argumentation framework with a *generalised notion of contrary*, as follows: $\mathcal{C} \subseteq 2^{\mathcal{A}} \times \mathcal{L}$. For every $(A, p) \in \mathcal{C}$, $A \cup \{p\}$ represents a set of sentences that cannot hold together, with A the designated "culprits" to be withdrawn should such a combination arise. Contraries as in standard assumption-based argumentation can be represented in terms of this generalised notion of contrary, by having a pair for each assumption, consisting of a singleton culprit set with that assumption and its contrary sentence. In the remainder, when clear from the context, we will refer to a generalised assumption-based argumentation framework simply as assumption-based argumentation framework.

In the assumption-based approach to argumentation, arguments are (forward or backward) deductions to conclusions, based upon assumptions.

Definition 1 (forward argument) *An argument $A \vdash_f p$ with conclusion p based on a set of assumptions A is a sequence $\beta_1, ..., \beta_m$ of sentences in $\mathcal{L}$, where $m > 0$ and $p = \beta_m$, such that, for all $i = 1, ..., m$: $\beta_i \in A$, or there exists $\frac{\alpha_1,\ldots,\alpha_n}{\beta_i} \in \mathcal{R}$ such that $\alpha_1, ..., \alpha_n \in \{\beta_1, ..., \beta_{i-1}\}$.*

Definition 2 (tight or backward argument) *Given a selection function, a tight argument $A \vdash_b p$ with conclusion p based on a set of assumptions A is a sequence of multi-sets $S_1, ..., S_m$, where $S_1 = \{p\}$, $S_m = A$, and for every $1 \leq i \leq m$, where σ is the selected sentence occurrence in S_i: 1) If σ is a non-assumption sentence then $S_{i+1} = S_i - \{\sigma\} \cup S$ for some inference rule of the form $\frac{S}{\sigma} \in \mathcal{R}$. 2) If σ is an assumption then $S_{i+1} = S_i$.*

Basically, tight arguments restrict the set A to include only assumptions that are relevant to the argument conclusion. Forward and backward arguments with the same conclusion are linked by

Theorem 1 ([Dung *et al.*, 2006]) *For every tight argument with conclusion p supported by A there exists an argument with conclusion p supported by A. For every argument with conclusion p supported by A and for every selection function, there exists a tight argument with conclusion p supported by some subset $A' \subseteq A$.*

In order to determine whether a conclusion (set of sentences) should be drawn, a set of assumptions needs to be identified providing an "acceptable" support for the conclusion. Various notions of "acceptable" support can be formalised [Dung, 1995; Bondarenko *et al.*, 1997; Dung *et al.*, 2002; 2006; 2007; Toni, 2007; Caminada, 2006], using a notion of attack amongst sets of assumptions. In this paper, for any sets of assumptions A and B we say that

Definition 3 (attack) *A attacks B if and only if there exists a pair $(P, q) \in \mathcal{C}$ such that $A \vdash_f q$ and $P \subseteq B$.*

We may transpose existing semantics for abstract [Dung, 1995] and assumption-based argumentation [Dung *et al.*, 2006] to the case of generalised assumption-based argumentation by saying that a set of assumptions A is

- *conflict-free* iff A does not attack itself
- *naive* iff A is maximally conflict-free
- *admissible* iff A is conflict-free and A attacks every set of assumptions B that attacks A
- *stable* iff A is conflict-free and attacks every set it does not include
- *semi-stable* iff A is complete where $\{A\} \cup \{B | A \ attacks \ B\}$ is maximal
- *preferred* iff A is maximally admissible
- *complete* iff A is admissible and includes every set B such that A attacks all sets attacking B (A defends B)
- *grounded* iff A is minimally complete
- *ideal* iff A is admissible and included in every preferred set.

The definitions given here are exactly the same as those used for standard assumption-based argumentation except for the stable and complete semantics which cannot be directly applied in generalised assumption-based argumentation and need a slight generalisation, and for the semi-stable semantics which has only been defined in the context of abstract argumentation [Caminada, 2006]. The new definitions for the stable and complete semantics collapse to the standard ones in every instance of a generalised framework where culprit sets of assumptions are singletons, as in standard assumption-based argumentation (cf. proof in appendix). The new definition for the semi-stable semantics is a possible adaptation of the original one to the case of generalised frameworks.

3. Transforming basic influence diagrams into argumentation frameworks

Basic influence diagrams can best be analysed using generalised assumption-based argumentation. Given a basic influence diagram, we define

- $\mathcal{L}$ as the set of all literals in the influence diagram
- $\mathcal{R}$ as the set of all inference rules of the form
  - $\frac{p_1,\ldots,p_n}{q}$ where "if p_1 and ... and p_n then q" is a dependency rule, or
  - $\frac{}{q}$ where "q" is a dependency rule
- $\mathcal{A}$ as the set of all atomic decisions and unobservable fundamental beliefs
- $\mathcal{C}$ as the set of all pairs (P, q) such that $(q, \neg q) \in \mathcal{L}^2$ and $P \vdash_b \neg q$.

Here, $\mathcal{L}$ represents the set of nodes of the diagram and $\mathcal{R}$ its arcs and dependency rules. $\mathcal{A}$ represents the uncertainties of the decision maker: he does not know which (atomic) decisions to make and which unobservable fundamental beliefs are true. $\mathcal{C}$ implicitly captures the conflicts between literals q and the relevant assumptions P supporting their negation $\neg q$. The notion of attack used in this paper is such that A attacks B if and only if there exists $(p, \neg p) \in \mathcal{L}^2$ such that $A \vdash_f p$ and $B \vdash_f \neg p$ (this can be easily proved using theorem 1). The assumption-based framework constructed captures in fact all the information originally contained in the basic influence diagram, except the positivity (+) and negativity (-) of goals which are not important for identifying all possible outcomes.

Since atomic decisions and unobservable fundamental beliefs have no incoming nodes, they are not conclusions of any inference rule. Therefore, the frameworks constructed are always guaranteed to be flat. Besides, these frameworks have the property of symmetry, which play a substantial role in the proofs given in this paper.

Property 1 (symmetry) *If A attacks B, then B attacks A.*

Proof 1 *If A attacks B, then there exists $(P, q) \in \mathcal{C}$ such that $P \subseteq B$ and $A \vdash_f q$. By definition of the contrary relation, $P \vdash_b \neg q$ is a tight argument. By definition, $\mathcal{C}$ also contains all pairs of the form $(Q, \neg q)$ where $Q \vdash_b q$ is a tight argument. Since $A \vdash_f q$ is an argument, there exist by theorem 1 a tight argument of the form $Q \vdash_b q$ such that $Q \subseteq A$. $P \vdash_b \neg q$ and $P \subseteq B$ so by theorem 1 one also has $B \vdash_f \neg q$. In summary, $(Q, \neg q) \in \mathcal{C}$, $Q \subseteq A$ and $B \vdash_f \neg q$, so B attacks A.*

In the remainder of the paper we show that the problem of finding the "rational" possible outcomes of decisions with respect to a given basic influence diagram may be reduced to the problem of finding "liberal stable" sets of assumptions in the corresponding argumentation framework.

4. Rationality for basic influence diagrams

When reasoning with rules, rationality is based on two requirements: consistency, or absence of conflicts, and closure under inference [Caminada and Amgoud, 2007; Toni, 2007]. These two properties are widely recognised as essential and we believe that a rational decision maker should conform to them. When using basic influence diagrams, it is however important to strengthen this basic notion of rationality by adding a third property. Basic influence diagrams allow the decision maker to express conflicts between beliefs in the form of contrary literals, such as dry and $\neg dry$ or $rain$ and $\neg rain$. Here, $\neg$ is meant to be used as classical negation, i.e. if p holds, then its contrary $\neg p$ does not hold, but if p does not hold, then its contrary $\neg p$ holds. So, for every pair $(p, \neg p) \in \mathcal{L}^2$, one must enforce the decision maker to choose to believe in either p or $\neg p$. We call this property *decidedness*. Some authors [Takahashi and Sawamura, 2004] insist on the philosophical importance of allowing descriptions of states of the world whereby both p and $\neg p$ hold, or even where neither of them holds. Such descriptions correspond for instance to logical formalisations of dialogues between parties sharing different views, but they are not meant to be taken as fully rational opinions of individuals. Therefore, we adopt the following

Definition 4 (rational outcome) *An outcome $O \subseteq \mathcal{L}$ is rational if and only if*

- $\forall (p, \neg p) \in \mathcal{L}^2$, *it is not the case that $p \in O$ and $\neg p \in O$ (consistency)*
- $\forall (p, \neg p) \in \mathcal{L}^2$, *$p \in O$ or $\neg p \in O$ (decidedness)*
- *there exists a set of assumptions A such that $O = \{p \in \mathcal{L} \mid A \vdash_f p\}$ (closure under dependency rules)*

Rational outcomes may contain positive and/or negative goals. When analysing a decision problem, it is indeed important to identify all possible outcomes, whether these are good or bad for the decision maker. Of course, the decision maker will try to achieve only the best ones, but ignoring bad possible outcomes would be quite imprudent.

Concerning the previous definition, note also that the closure property is expressed in terms of the assumption-based framework on which dependency rules and influence diagrams are mapped. Indeed, the notion $\vdash_f$ there corresponds exactly to the notion of reasoning with dependency rules we are after. Finally, it can be remarked that the set A is always unique and given by $A = O \cap \mathcal{A}$. In this paper, rational opinions of the decision maker correspond to rational outcomes.

Definition 5 (rational opinion) *A rational opinion is a set of assumptions A such that the set $O(A)$ defined as $\{p \in \mathcal{L} \mid A \vdash_f p\}$ is a rational outcome.*

Identifying all rational outcomes is equivalent to identifying all rational opinions. In the next section, we introduce a new argumentation semantics that allows to characterise rational opinions.

5. Liberal stable semantics

In general, the conflict-freeness of A is sufficient to guarantee the consistency of $O(A)$.

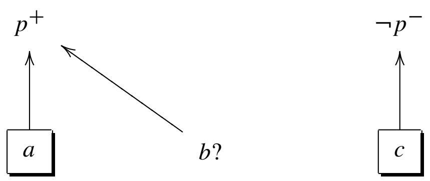
Lemma 1 (consistency) *$\nexists p \in \mathcal{L}$ such that $p \in O(A)$ and $\neg p \in O(A)$ iff A does not attack itself.*

Proof 2 *It is equivalent to prove that $\exists p \in \mathcal{L}$ such that $p \in O(A)$ and $\neg p \in O(A)$ iff A attacks itself. By definition of an attack, A attacks itself iff $\exists (P, q) \in \mathcal{C}$ such that $P \subseteq A$ and $A \vdash_f q$. By definition of the contrary relation, A attacks itself iff there exists a tight argument $P \vdash_b \neg q$ such that $P \subseteq A$ and $A \vdash_f q$. By theorem 1, this is also equivalent to saying that A attacks itself iff there exists $q \in \mathcal{L}$ such that $A \vdash_f \neg q$ and $A \vdash_f q$.*

Unfortunately, the semantics mentioned in the previous section fail to characterise the rationality of opinions.

Theorem 2 *None of the notions of acceptability in section 2 are such that $O(A)$ is a rational outcome if and only if A is acceptable.*

Proof 3 *Let us consider the following basic influence diagram and influence rules*



if a and b then p
if c then ¬p

We obtain a generalised assumption-based framework with $\mathcal{L} = \{a, b, c, p, \neg p\}$*,* $\mathcal{R} = \{\frac{a,b}{p}, \frac{c}{\neg p}\}$*,* $\mathcal{A} = \{a, b, c\}$ *and* $\mathcal{C} = \{(\{a, b\}, \neg p), (\{c\}, p)\}$*. The rational opinions are* $\{c\}$*,* $\{a, b\}$*,* $\{a, c\}$ *and* $\{b, c\}$*.* $\{\}$ *is conflict-free but is not rational.* $\{c\}$ *and* $\{a, c\}$ *cannot be both naive but are nevertheless both rational.* $\{\}$ *is admissible but is not rational.* $\{c\}$ *is not stable but is nevertheless rational.* $\{c\}$ *and* $\{a, c\}$ *cannot both be preferred but are nevertheless both rational.* $\{c\}$ *and* $\{a, c\}$ *cannot both be grounded but are nevertheless both rational.* $\{\}$ *is ideal but is not rational.* $\{c\}$ *is not complete (it defends* $\{a\}$ *(since* $\{a\}$ *is not attacked) and does not contain a) but is nevertheless rational.* $\{c\}$ *is not complete and a fortiori not semi-stable, but is nevertheless rational.*

This justifies the need for a new semantics. For any $A \subseteq \mathcal{A}$, let us denote $att(A) = \{B \subseteq \mathcal{A} \mid A \text{ attacks } B\}$. We define

Definition 6 (liberal stability) *A is liberal stable if and only if*

- *A is conflict-free, and*
- *there is no conflict-free set B such that* $att(B) \supset att(A)$*.*

For an abstract argumentation framework [Dung, 1995] with set of arguments Arg and attack relation $R \subseteq Arg \times Arg$, liberal stability can be defined as follows. If we denote $att(S) = \{B \in Arg \mid \exists A \in S, A \; R \; B\}$ for any $S \subseteq Arg$), then the set of arguments $S \subseteq Arg$ is liberal stable if and only if S is conflict-free and there is no conflict-free set $X \subseteq Arg$ such that $att(X) \supset att(A)$.

This new semantics enables to exactly identify the rational opinions in the example of proof 3. Indeed, note first that the conflict-free sets of assumptions are the strict subsets of $\mathcal{A}$. It is clear that if a set of assumptions A is such that both $A \not\vdash_f p$ and $A \not\vdash_f \neg p$ then $att(A) = \emptyset$. Such a set is not liberal stable as for instance with the conflict-free set $B = \{c\}$, we have $att(B) = \{\{a, b\}, \{a, b, c\}\} \supset \emptyset$. So, it is necessary for a liberal stable set to be such that $A \vdash_f p$ or $A \vdash_f \neg p$. The only conflict-free set such that $A \vdash_f p$ is $\{a, b\}$ and the only conflict-free sets such that $A \vdash_f \neg p$ are $\{c\}$, $\{a, c\}$ and $\{b, c\}$. For all these sets, the only B such that $att(B) \supset att(A)$ is $B = \mathcal{A}$ which is not conflict-free. This proves that these sets are the liberal stable ones and as we have seen before they exactly correspond to the rational opinions. We prove more generally that this holds whenever the argumentation framework has the property of

Hypothesis 1 (extensibility) *Every conflict-free set is contained in at least one rational opinion.*

Extensibility holds in the previous example: the conflict-free sets are the strict subsets of $\mathcal{A}$, so either they do not contain a and are included in the rational set $\{b, c\}$, or they do not contain b and are included in the rational set $\{a, c\}$, or they do not contain c and are included in the rational set $\{a, b\}$. The main result of this paper is the following

Theorem 3 *i) If A is rational then A is liberal stable. Under extensibility, it also holds that ii) if A is liberal stable then A is rational.*

Proof 4 *According to the consistency lemma 1 and the definition of rational outcome, if A is rational then $O(A)$ is consistent and A is conflict-free. Conversely, if A is liberal stable then A is conflict-free and $O(A)$ is consistent. Consequently, in the remainder of the proof, in i) it only remains to prove that there is no conflict-free set attacking a strict superset of $att(A)$ and in ii) that for all $(p, \neg p) \in \mathcal{L}^2$, $p \in O(A)$ or $\neg p \in O(A)$.*

Proof of i): Let A be a rational set. Assume that there is a conflict-free set B such that $att(B) \supset att(A)$. Then $att(B) - att(A) \neq \emptyset$, i.e. there exists $C \in att(B) - att(A)$. So, there exist $(p, \neg p) \in \mathcal{L}^2$ and a tight argument of the form $P \vdash_b \neg p$ such that: $(P, p) \in \mathcal{C}$, $P \subseteq C$ and $B \vdash_f p$. Since A does not attack C, we also have $A \nvdash_f p$. A is rational and therefore $A \vdash_f \neg p$ or $A \vdash_f p$. We know that $A \nvdash_f p$ so $A \vdash_f \neg p$. Since $B \vdash_f p$ and by theorem 1, we can find a tight argument $Q \vdash_b p$ where $Q \subseteq B$. By definition of $\mathcal{C}$: $(Q, \neg p) \in \mathcal{C}$. In summary, we have $(Q, \neg p) \in \mathcal{C}$, $Q \subseteq B$ and $A \vdash_f \neg p$ so A attacks B. $B \in att(A)$ and $att(B) \supset att(A)$ implies $B \in att(B)$, which is absurd because B is conflict-free.

Proof of ii): Let A be a liberal stable set and $(p, \neg p) \in \mathcal{L}^2$. Assume that $p \notin O(A)$ and $\neg p \notin O(A)$. A is liberal stable and a fortiori conflict-free. Since we have assumed that the framework is extensible, A can be extended to a set $A' \supseteq A$ that is rational. A' is therefore conflict-free and notably such that $A' \vdash_f p$ or $A' \vdash_f \neg p$. We may assume without loss of generality that $A' \vdash_f p$. In an assumption-based framework drawn from a basic influence diagram, every sentence of $\mathcal{L}$ is supported by at least one (tight) argument. Let then $Q \vdash_b \neg p$ be a tight argument supporting $\neg p$. Clearly, $A' \supseteq A$ implies that $att(A') \supseteq att(A)$. Note that in fact A' is conflict-free and $att(A') \supset att(A)$ (since A' attacks Q but A does not attack it). This is absurd because A is liberal stable.

Fortunately, extensibility holds whenever naive opinions are decided. Extensibility is therefore a quite natural property of frameworks derived from basic influence diagrams.

Property 2 *i) $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$ is extensible if and only if ii) for every naive set N of assumptions, the opinion $O(N)$ is decided.*

Proof 5 *i) $\Rightarrow$ ii) Let N be a naive set. Assume $O(N)$ is not decided. Then by conflict-freeness of N and i), N can be extended to a larger set $N' \supseteq N$ that is rational. N' is rational and a fortiori consistent, i.e conflict-free by the consistency lemma 1. $O(N')$ is decided, and therefore it is clear that $N' \supset N$. N' is a strict superset of N and is conflict-free, which contradicts the naiveness of N.*

ii) $\Rightarrow$ i) Let A be a conflict-free set. Then A is contained in a maximally conflict-free (naive) set N. N is rational since $O(N)$ satisfies the property of decidedness (by ii) and the property of consistency by the consistency lemma 1.

In the umbrella example, the naive sets and their corresponding opinions are

- $N_1 = \{umbrella, clouds\}$, $O(N_1) = \{umbrella, clouds, loaded, dry\}$,
- $N_2 = \{umbrella, \neg clouds\}$, $O(N_2) = \{umbrella, \neg clouds, loaded, dry\}$,
- $N_3 = \{\neg umbrella, \neg clouds\}$, $O(N_3) = \{\neg umbrella, \neg clouds, \neg loaded, dry\}$,
- $N_4 = \{\neg umbrella, clouds\}$, $O(N_4) = \{\neg umbrella, clouds, \neg loaded, \neg dry\}$.

The four opinions are decidedness and the framework is therefore extensible. In fact, the naive sets here correspond exactly to the liberal stable sets. There are therefore four possible outcomes and these solutions are those we would intuitively expect.

Frameworks derived from basic influence diagrams may not be extensible. For instance, a framework is not extensible when it admits at least one conflict-free set but does not admit any rational set/possible outcome. This happens e.g. with $\mathcal{L} = \{d, p, \neg p\}$ $\mathcal{R} = \{\frac{d}{p}, \frac{p}{\neg p}\}$, $\mathcal{A} = \{d\}$, $\mathcal{C} = \{(\{d\}, p), (\{d\}, \neg p)\}$. Such frameworks/diagrams are "pathological" and do not conform to the idea that in the practical domain of decision, either p or $\neg p$ holds. Frameworks drawn from diagrams are not extensible either when they do not admit any conflict-free sets of assumptions. This occurs when it is possible to infere both p and $\neg p$ simply from facts of the diagram. Such frameworks are inconsistent and all semantics including the liberal stable semantics are then anyway empty. Extensibility is a stronger notion than consistency, i.e. whenever a framework is extensible, it is also consistent. In general, one has

Property 3 *The existence of conflict-free sets implies the existence of liberal stable ones.*

Note that there may not always exist a conflict-free sets of assumption. For instance, if in the basic influence diagram q and $\neg q$ are facts, then in the corresponding framework, {} attacks itself, and as a matter of consequence, all sets of assumptions also attack themselves. The following theorem links the semantics of liberal stability to the others.

Theorem 4 *Every stable set is liberal stable and every liberal stable set is conflict-free and admissible. If extensibility holds, then every naive, stable or preferred set is liberal stable and every liberal stable set is conflict-free and admissible (cf. figure 2).*

Semantics s	$s \Rightarrow$ **liberal stable ?**	**liberal stable** $\Rightarrow s$ **?**
conflict-free	$\times^*$	$\surd^0$
naive	E^1	$\times^*$
admissible	$\times^*$	$\surd^2$
stable	$\surd^3$	$\times^*$
semi-stable	$\times^4$	$\times^*$
preferred	E^5	$\times^*$
complete	$\times^6$	$\times^*$
grounded	$\times^7$	$\times^*$
ideal	$\times^*$	$\times^8$

Figure 2. Links between liberal stability and existing semantics for argumentation. $\surd$ denotes an implication that holds in general, E an implication that holds under extensibility and $\times$ an implication that does not generally hold, even under the extensibility hypothesis. The symbols in exponent indicate the paragraph in which the result is proved.

Proof 6 *All the results marked with * directly follow from the proof of theorem 2. We now denote* $cont(A) = \{p \in \mathcal{L} | A \vdash_f p \text{ and } (p, \neg p) \in \mathcal{L}^2\}$.

0) Every liberal stable set is conflict-free by definition.

1) Let A be a naive set. A is conflict-free. Assume there exists a conflict-free set B such that $att(B) \supset att(A)$. *Then, we would have* $cont(B) \supset cont(A)$. *There would then exist* $p \in$

$cont(B) - cont(A)$. $A \vdash_f \neg p$ otherwise A could be extended by extensibility to a conflict-free set that entails p or $\neg p$ and that set would be strictly larger than A. This would contradict the maximality of A. So, $\neg p \in cont(A)$. Since $cont(B) \supset cont(A)$, $\neg p \in cont(B)$. Since $B \vdash_f p$, there exists by theorem 1 a tight argument $P \vdash_b p$ with $P \subseteq B$. We have $(P, \neg p) \in \mathcal{C}$, $P \subseteq B$ and $B \vdash_f \neg p$: B attacks itself, which is absurd.

2) In symmetric abstract frameworks, admissibility collapses with conflict-freeness [Coste-Marquis et al.*, 2005]. The same thing happens in symmetric frameworks derived from basic influence diagrams. Since liberal stability implies conflict-freeness, it is also clear that liberal stability implies admissibility.*

3) Let A be a stable set. A attacks all the sets that it is possible to attack without attacking itself, so A is definitely liberal stable.

5) Preferred sets are defined as maximally admissible and by symmetry actually coincide with the maximally conflict-free (i.e. naive) ones. The proof of 1) allows us to conclude that preferred sets are liberal stable.

6) Consider the simple case where $\mathcal{L} = \{a, b, x, p, \neg p\}$, $\mathcal{R} = \{\frac{x}{p}, \frac{b}{\neg p}, \frac{a,b}{\neg p}\}$, $\mathcal{A} = \{a, b, x\}$ and $\mathcal{C} = \{(\{x\}, \{\neg p\}), (\{b\}, \{p\}), (\{a, b\}, \{p\})\}$. In this case, $A = \{a\}$ is complete but is not liberal stable.

7) In the same framework as in 6), note that all complete sets must include $\{a\}$ (which is not attacked) and therefore $A = \{a\}$ is minimally complete (grounded). However, A is not liberal stable.

4) Semi-stable sets must be complete. In the same framework as in 6), all complete sets contain a. So, $\{x\}$ is neither complete nor semi-stable. However, $\{x\}$ is liberal stable.

8) In the case where $\mathcal{L} = \{a, b, c, p, \neg p\}$, $\mathcal{R} = \{\frac{a,b}{p}, \frac{c}{\neg p}\}$, $\mathcal{A} = \{a, b, c\}$ and $\mathcal{C} = \{(\{a, b\}, \neg p), (\{c\}, p)\}$, the preferred sets collapse to the naive ones by symmetry, so the preferred sets are $\{a, b\}$, $\{c, a\}$ and $\{c, b\}$. Ideal sets must be admissible and contained in the intersection $\{a, b\} \cap \{c, a\} \cap \{c, b\} = \emptyset$. So, the only ideal set is $\{\}$ but none of the liberal stable sets is empty.

6. Summary and future work

Practical situations of decision making under strict uncertainty may be described qualitatively using basic influence diagrams. These diagrams give a logical structure to the decision domain and reveal the decision maker's most fundamental uncertainties. Basic influence diagrams can best be analysed and resolved using assumption-based argumentation. In the argumentation framework derived from a diagram and under a quite natural hypothesis called extensibility, we have proved that the possible outcomes of decisions are in one-to-one correspondence with the consequences of liberal stable sets of assumptions. Still under extensibility, we have shown that the set of liberal stable sets of assumptions includes all the naive, stable and preferred sets of assumptions and that it was included in the set of admissible sets of assumptions. In future work, we intend to develop an algorithm for the computation of liberal stable sets of assumptions and use this algorithm to transform basic influence diagrams into proper decision tables.

Acknowledgements

This work was partially funded by the Sixth Framework IST programme of the European Commission, under the 035200 ARGUGRID project. The second author has also been supported by a UK Royal Academy of Engineering/Leverhulme Trust senior fellowship.

References

[Amgoud and Prade, 2004] L. Amgoud and H. Prade. Using arguments for making decisions: A possibilistic logic approach. In *Proceedings 20th Conference of Uncertainty in AI*, pages 10–17, 2004.

[Bondarenko *et al.*, 1997] A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93(1–2):63–101, 1997.

[Caminada and Amgoud, 2007] M. Caminada and L. Amgoud. An axiomatic account of formal argumentation. *Artificial Intelligence*, 2007.

[Caminada, 2006] M. Caminada. Semi-stable semantics. In *Proceedings of 1st International Conference on Computational Models of Argument*, 2006.

[Coste-Marquis *et al.*, 2005] C. Coste-Marquis, C. Devred, and P. Marquis. Symmetric argumentation frameworks. In *Proceedings 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 317–328, 2005.

[Dung *et al.*, 2002] P. M. Dung, P. Mancarella, and F. Toni. Argumentation-based proof procedures for credulous and sceptical non-monotonic reasoning. In *Computational Logic: Logic Programming and Beyond, Essays in Honour of R. A. Kowalski, Part II*, pages 289–310. Springer, 2002.

[Dung *et al.*, 2006] P. M. Dung, R. A. Kowalski, and F. Toni. Dialectic proof procedures for assumption-based, admissible argumentation. *Artificial Intelligence*, 170(2):114–159, 2006.

[Dung *et al.*, 2007] P.M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence, Special Issue on Argumentation in Artificial Intelligence*, 171(10–15):642–674, 2007.

[Dung, 1995] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. *Artificial Intelligence*, 77(2):321–257, 1995.

[French, 1987] S. French. *Decision theory: an introduction to the mathematics of rationality*. Ellis Horwood, 1987.

[Howard and Matheson, 1981] R. A. Howard and J. E. Matheson. Influence diagrams. In *Readings on the Principles and Applications of Decision Analysis*, volume II, pages 721–762. 1981.

[Morge and Mancarella, 2007] M. Morge and P. Mancarella. The hedgehog and the fox. An argumentation-based decision support system. In *Proceedings 4th International Workshop on Argumentation in Multi-Agent Systems*, 2007.

[Pearl, 1986] J. Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288, 1986.

[Takahashi and Sawamura, 2004] T. Takahashi and H. Sawamura. A logic of multiple-valued argumentation. In *Proceedings 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems*, pages 800–807, 2004.

[Toni, 2007] F. Toni. Assumption-based argumentation for closed and consistent defeasible reasoning. In *Proceedings 1st International Workshop on Juris-informatics in association with the 21st Annual Conference of The Japanese Society for Artificial Intelligence*, 2007.

Appendix

Lemma 2 *Given a standard assumption-based argumentation framework and a set A of assumptions*

1. *A is stable in the standard sense (see [Bondarenko* et al.*, 1997]) iff A is stable in the sense of section 2*
2. *A is complete in the standard sense (see [Bondarenko* et al.*, 1997]) iff A is complete in the sense of section 2*

Proof 7 *1a) If A is stable in the standard sense, then A is conflict-free. If $\neg(A \supseteq B)$ then $\exists b \in B - A$. $b \notin A$, A attacks $\{b\}$ and therefore A attacks B. Hence, A is stable in the new sense. 1b) If A is stable in the new sense, then A is conflict-free. If $x \notin A$ then $\neg(A \supseteq \{x\})$ so A attacks $\{x\}$. A is stable in the standard sense. 2a) If A is complete in the standard sense, then A is admissible. Assume B is defended by A and let $b \in B$. If $\{b\}$ is not attacked, then A defends $\{b\}$ and therefore $b \in A$. Otherwise, $\{b\}$ is attacked. A defends B so A defends $\{b\}$ and by completeness $b \in A$. So, $A \supseteq B$. A is complete in the new sense. 2b) If A is complete in the new sense, then A is admissible. If A defends $\{x\}$, then $A \supseteq \{x\}$, i.e. $x \in A$. A is complete in the standard sense.*