Preferential Reasoning Based On Abstract Argumentation Semantics

Ofer ARIELI [a] and Tjitze REINSTRA [b]
[a] *School of Computer Science, The Academic College of Tel-Aviv, Israel*
[b] *Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg*

Abstract. We introduce a preferential-based setting for reasoning with different types of argumentation-based semantics, including those that are not necessarily conflict-free or admissible. The induced entailments are defined by n-valued labeling and may be computed by answer-set programs.

Keywords. semantics for abstract argumentation, preferential reasoning, n-valued labeling, answer-set programming.

1. Introduction

The seminal paper of Dung [12], published nearly twenty years ago, gave rise to a wide variety of graph-based approaches for representing disputes and giving semantics to argumentation-based conflicts (see, e.g., the surveys in [5,15]). Generally, these approaches indicate what sets of arguments can be collectively accepted and what can be inferred based on these sets. Often, the sets of arguments are assumed to have some basic properties and are obtained by certain preference criteria posed on the set of all arguments.

The idea of this paper is to capture the common background behind many of the semantic approaches to abstract argumentation in a formal and uniform way. This will allow us to provide a general method of representing different kinds of argumentation semantics, including those that do not presuppose the 'standard' assumptions on acceptable arguments (like conflict-freeness and admissibility), or those that include 'extra' assumptions on the acceptable arguments in the form of integrity constraints. Another important aspect of our approach is its applicability in terms of answer-set programs.

The rest of this paper is organized as follows: in the next section we briefly review some basic concepts behind abstract argumentation and introduce our approach for representing their semantics. In Section 3 we apply our method for different cases and then exemplify this in Section 4. In Section 5 we show how our approach is implementable by ASP, and in Section 6 we conclude.

2. Preferential n-Labeling Semantics

Definition 1 An *argumentation framework* [12] is a pair $\mathcal{AF} = \langle Args, Attack \rangle$, where $Args$ is an enumerable set of elements, called *arguments*, and $Attack$ is a relation on

$Args \times Args$ whose instances are called *attacks*. When $(A, B) \in Attack$ we say that A *attacks* B (or that B is *attacked by* A).

In the sequel we denote by A^+ the set of arguments attacked by A, that is, $A^+ = \{B \in Args \mid Attack(A, B)\}$, and denote by A^- the set of arguments that attack A, i.e., $A^- = \{B \in Args \mid Attack(B, A)\}$.

Our approach extends the standard 3-valued labeling semantics for abstract argumentation framework [9,17] by allowing an arbitrary number of labels as follows.

Definition 2 Given an argumentation framework $\mathcal{AF} = \langle Args, Attack\rangle$, an *$n$-valued labeling* of $\mathcal{AF}$ is a complete function $lab\colon Args \to \{\mathsf{val}_1, \ldots, \mathsf{val}_n\}$. For $1 \le i \le n$, we denote $\mathsf{Val}_i(lab) = \{A \in Args \mid lab(A) = \mathsf{val}_i\}$. The set of all the n-valued labelings on $Args$ is denoted $\mathsf{Lab}^n(Args)$.

To simplify the reading, the labels of 2-valued labelings will be denoted in and out. For 3-valued labelings we shall use in addition the label undec, and for 4-valued labelings we shall replace undec by none and add the label both. The corresponding sets of arguments will be denoted similarly (e.g., $\mathsf{In}(lab) = \{A \in Args \mid lab(A) = \mathsf{in}\}$).

The semantics of an argumentation framework is determined by first focusing on the n-valued labelings that satisfy certain conditions (specified by the propositional formulas in $\mathsf{Cond}(\mathcal{AF})$, see below), and then considering only the most preferred labelings (with respect to some normality considerations, represented by $\le$). This is formalized next.

Given an argumentation framework $\mathcal{AF} = \langle Args, Attack\rangle$ where $Args = \{A_1, A_2, \ldots\}$. For expressing the condition(s) that the 'legitimate' n-valued labelings of $\mathcal{AF}$ should meet, we fix a propositional language $\mathcal{L}^n_{Args}$, whose atomic formulas are associated with the labeling of the argument of $\mathcal{AF}$, and whose compound formulas are generated by the following BNF:

$$
\begin{array}{lcl}
A \in Arguments & := & A_1 \mid A_2 \mid \ldots \\
\mathsf{v} \in LabelValues & := & \mathsf{v}_1 \mid \ldots \mid \mathsf{v}_n \\
\psi, \phi \in Formulas & := & \mathsf{val}(A, \mathsf{v}) \mid \neg\psi \mid \psi \vee \phi \mid \psi \wedge \phi \mid \psi \supset \phi \mid \mathsf{f}
\end{array}
$$

For defining the semantics of $\mathcal{L}^n_{Args}$ we identify the n-valued labelings functions with n-valued valuations for formulas in $\mathcal{L}^n_{Args}$. Accordingly, the satisfaction relation is defined as follows:

Definition 3 The *satisfaction relation* for $\mathcal{L}^n_{Args}$ is a binary relation $\models^n_{Args}$ between n-valued labelings of $Args$ and formulas of $\mathcal{L}^n_{Args}$, defined as follows:

- $lab \models^n_{Args} \mathsf{val}(A, \mathsf{v})$ iff $lab(A) = \mathsf{v}$,
- $lab \models^n_{Args} \neg\psi$ iff $lab \not\models^n_{Args} \psi$,
- $lab \models^n_{Args} \psi \vee \phi$ iff $lab \models^n_{Args} \psi$ or $lab \models^n_{Args} \phi$,
- $lab \models^n_{Args} \psi \wedge \phi$ iff $lab \models^n_{Args} \psi$ and $lab \models^n_{Args} \phi$,
- $lab \models^n_{Args} \psi \supset \phi$ iff $lab \not\models^n_{Args} \psi$ or $lab \models^n_{Args} \phi$,
- $lab \not\models^n_{Args} \mathsf{f}$.

We denote: $mod^n_{Args}(\psi) = \{lab \in \mathsf{Lab}^n(Args) \mid lab \models^n_{Args} \psi\}$ and $mod^n_{Args}(\Gamma) = \bigcap_{\psi\in\Gamma} mod^n_{Args}(\psi)$. When $Args$ and n are known and fixed, we shall sometimes omit them from these notations.

Preferential n-labeling semantics is now defined as follows:

Definition 4 Let $\mathcal{AF} = \langle Args, Attack \rangle$ be an argumentation framework and $\mathcal{L}^n_{Args}$ a corresponding propositional language for some $n \in \mathbb{N}$. A *preferential n-labeling semantics* for $\mathcal{AF}$ is a triple $\mathcal{SEM}(\mathcal{AF}) = \langle \mathsf{Lab}^n(Args), \mathsf{Cond}(\mathcal{AF}), < \rangle$, where $\mathsf{Cond}(\mathcal{AF})$ is a set of formulas in $\mathcal{L}^n_{Args}$ and $<$ is a relation on $\mathsf{Lab}^n(Args) \times \mathsf{Lab}^n(Args)$.

A preferential n-labeling semantics thus consists of the possible labelings for $\mathcal{AF}$, conditions that express what the valid labelings are, and a preference criterion for choosing the 'best' labelings among the valid ones. The induced conclusions are defined next.

Definition 5 Let $\mathcal{SEM}(\mathcal{AF}) = \langle \mathsf{Lab}^n(Args), \mathsf{Cond}(\mathcal{AF}), < \rangle$ be a preferential n-labeling semantics for an argumentation framework $\mathcal{AF} = \langle Args, Attack \rangle$.

- The *$\mathcal{SEM}$-labelings of $\mathcal{AF}$* are the elements of the following set: $mod^n_<(\mathsf{Cond}(\mathcal{AF})) = \{lab \in mod^n_{Args}(\mathsf{Cond}(\mathcal{AF})) \mid$ $\neg\exists\, lab' \in mod^n_{Args}(\mathsf{Cond}(\mathcal{AF}))$ such that $lab' < lab\}$.
- ψ is a *skeptical $\mathcal{SEM}$-conclusion* of $\mathcal{AF}$, if $mod^n_<(\mathsf{Cond}(\mathcal{AF})) \subseteq mod^n_{Args}(\psi)$.[1]
- ψ is a *credulous $\mathcal{SEM}$-conclusion*, if $mod^n_<(\mathsf{Cond}(\mathcal{AF})) \cap mod^n_{Args}(\psi) \neq \emptyset$.

3. Applications

In this section we show that the preferential semantics described in the previous section captures different forms of reasoning with abstract argumentation frameworks. We start with Dung-type semantics [5,12], in which the underlying labelings are both conflict-free and admissible (Section 3.1). We also show how integrity constraints may be handled in this context (Section 3.2). Then we describe semantics in which either of these conditions is relaxed: fallback semantics [8] and Jakobovits-Vermeir's semantics [14], which are coherent-based approaches that preserve conflict-freeness but give-up admissibility (Sections 3.3 and 3.5), and conflict-tolerant semantics [1,3], which is a paraconsistent approach that sticks to admissibility but abandons conflict-freeness (Section 3.4).

3.1. Standard Semantics Based On Complete Labelings

Many of the semantical approaches to reasoning with Dung-style abstract argumentation frameworks can be represented by 3-valued semantics as described in Definition 4. Some of them are considered next.

Definition 6 Given an argumentation framework $\mathcal{AF} = \langle Args, Attack \rangle$, we consider the following conditions on a 3-valued labeling lab of $\mathcal{AF}$:

Pos: If $lab(A) = \mathsf{in}$ then for every $B \in A^-$ it holds that $lab(B) = \mathsf{out}$.
Neg: If $lab(A) = \mathsf{out}$ then there exists some $B \in A^-$ such that $lab(B) = \mathsf{in}$.
Neither1: If $lab(A) = \mathsf{undec}$ then there is no $B \in A^-$ such that $lab(B) = \mathsf{in}$.
Neither2: If $lab(A) = \mathsf{undec}$ then there is $B \in A^-$ such that $lab(B) \neq \mathsf{out}$.

[1] Intuitively, this means that all the $<$-most preferred valid labelings of $\mathcal{AF}$ are models of ψ.

Now, given a 3-valued labeling lab of $\mathcal{AF}$, we say that

1. lab is *conflict-free* if it satisfies conditions **Pos** and **Neither1**.[2]
2. lab is *admissible* if it is conflict-free and satisfies condition **Neg**.
3. lab is *complete* if it is admissible and satisfies condition **Neither2**.

Let lab_{cmp} be a complete 3-valued labeling of $\mathcal{AF}$. Below, the minimum and maximum are taken with respect to set inclusion.

- lab_{cmp} is a *grounded labeling* of $\mathcal{AF}$ iff $\mathsf{In}(lab_{cmp})$ is minimal in the set $\{\mathsf{In}(lab) \mid lab$ is a 3-valued complete labeling of $\mathcal{AF}\}$.
- lab_{cmp} is a *preferred labeling* of $\mathcal{AF}$ iff $\mathsf{In}(lab_{cmp})$ is maximal in the set $\{\mathsf{In}(lab) \mid lab$ is a 3-valued complete labeling of $\mathcal{AF}\}$.
- lab_{cmp} is a *stable labeling* of $\mathcal{AF}$ iff $\mathsf{Undec}(lab_{cmp}) = \emptyset$.
- lab_{cmp} is a *semi-stable labeling* of $\mathcal{AF}$ iff $\mathsf{Undec}(lab_{cmp})$ is minimal in $\{\mathsf{Undec}(lab) \mid lab$ is a 3-valued complete labeling of $\mathcal{AF}\}$.

Note 7 Intuitively, a labeling is conflict-free if all the neighbors of an in-labeled argument are labeled out. This definition of a conflict-free labeling is equivalent to what is called a *subcomplete* labeling in [7] and it deviates from the definitions of a conflict-free labeling used, e.g., in [9], where a conflict-free labeling should satisfy Condition **Neg** and the following weaker version of **Pos**:

w-Pos If $lab(A) = \mathsf{in}$ then there is no $B \in A^-$ such that $lab(B) = \mathsf{in}$.

The reason for using a modified definition is that it assures the following two properties that are not guaranteed by **Neg** and **w-Pos**:

1. A partial labeling of a conflict-free labeling (obtained by removing arguments from the original labeling) is still conflict-free:

   Proposition 8 *For every conflict-free labeling lab of $\mathcal{AF} = \langle Args, Attack \rangle$ and every restriction $\mathcal{AF}' = \langle Args', Attack \cap (Args' \times Args') \rangle$ of $\mathcal{AF}$ to a subset $Args' \subseteq Args$, the restriction of lab to $Args'$ is a conflict-free labeling of $\mathcal{AF}'$.*

   Proof. Simple, and omitted due to short of space. □

2. A conflict-free labeling can be turned into (part of) a complete labeling by adding arguments and/or attacks:

   Proposition 9 *For every conflict-free labeling lab of $\mathcal{AF} = \langle Args, Attack \rangle$, there is an expansion $\mathcal{AF}' = \langle Args', Attack' \rangle$ of $\mathcal{AF}$ for some $Args \subseteq Args'$ and $Attack \subseteq Attack'$ such that lab is the restriction to Args of a complete labeling lab' of $\mathcal{AF}'$.*

   Proof. (Sketch) Let lab be a conflict-free labeling of $\mathcal{AF} = \langle Args, Attack \rangle$, and let $\mathsf{Illegal}(lab)$ be the subset of $Args$, defined as follows:

   Definition 10 *The set $\mathsf{Illegal}(lab)$ (of arguments that have illegal labelings according to complete semantics) consists of the arguments $A \in Args$ that satisfy either of the following conditions:*

[2] Here and in what follows satisfaction of the conditions is taken with respect to every $A \in Args$.

- $lab(A) = \mathsf{out}$ *and there is no* $B \in A^-$ *such that* $lab(A) = \mathsf{in}$*,*
- $lab(A) = \mathsf{undec}$ *and either there is some* $B \in A^-$ *such that* $lab(B) = \mathsf{in}$ *or there is no* $B \in A^-$ *such that* $lab(B) = \mathsf{undec}$*.*

Consider now the argumentation framework $\mathcal{AF}' = \langle Args \cup \{\alpha, \beta\}, Attack \cup Attack' \rangle$, where $Attack' = \{(\beta, \beta)\} \cup \{(\alpha, A) \mid lab(A) = \mathsf{out}, A \in \mathsf{Illegal}(lab)\} \cup \{(\beta, A) \mid lab(A) = \mathsf{undec}, A \in \mathsf{Illegal}(lab)\}$. We define lab' by $lab'(\alpha) = \mathsf{in}$, $lab'(\beta) = \mathsf{undec}$ and $lab'(A) = lab(A)$ if $A \in Args$. It can be checked that lab' is a complete labeling of $\mathcal{AF}'$ and that the restriction of lab' to $Args$ equals lab. □

Additionally, our notion of admissible labeling is not the same as that of [9], where admissible labelings are defined by Conditions **Pos** and **Neg**. The reason is that in our case there is a one-to-one correspondence between admissible labelings and admissible extensions, while this is not the case according to [9].

We note, finally, that complete labelings in our sense are the same as complete labelings in the sense of [9] (and so are grounded, preferred, stable, and semi-stable labelings). In [9], Conditions **Neither1** and **Neither2** are merged to one property:

Neither If $lab(A) = \mathsf{undec}$ then not for every $B \in A^-$ it holds that $lab(B) = \mathsf{out}$ and there does not exist a $B \in A^-$ such that $lab(B) = \mathsf{in}$.

Definition 11 Let $\mathcal{AF} = \langle Args, Attack \rangle$ and let $A \in Args$.

- A *skeptically follows* from $\mathcal{AF}$ by complete (resp, grounded, preferred, stable, semi-stable) semantics, if $lab(A) = \mathsf{in}$ for *every* complete (resp. grounded, preferred, stable, semi-stable) 3-valued labeling lab of $\mathcal{AF}$.
- A *credulously follows* from $\mathcal{AF}$ by complete (resp, grounded, preferred, stable, semi-stable) semantics, if $lab(A) = \mathsf{in}$ for *some* complete (resp. grounded, preferred, stable, semi-stable) 3-valued labeling lab of $\mathcal{AF}$.

Next we show that the argumentation semantics in Definition 4 indeed capture those in Definition 6.

Definition 12 Let $\mathcal{AF} = \langle Args, Attack \rangle$ be an argumentation framework. Consider the following theory in the language $\mathcal{L}^3_{Args}$ for 3-states argumentation:

$$\mathsf{CMP}(\mathcal{AF}) = \bigcup_{A \in Args} \left\{ \begin{array}{l} \mathsf{val}(A, \mathsf{in}) \supset \bigwedge_{B \in A^-} \mathsf{val}(B, \mathsf{out}), \\ \mathsf{val}(A, \mathsf{out}) \supset \bigvee_{B \in A^-} \mathsf{val}(B, \mathsf{in}), \\ \mathsf{val}(A, \mathsf{undec}) \supset \left(\neg \bigwedge_{B \in A^-} \mathsf{val}(B, \mathsf{out}) \wedge \neg \bigvee_{B \in A^-} \mathsf{val}(B, \mathsf{in})\right) \end{array} \right\}$$

Definition 13 Let $\mathcal{AF} = \langle Args, Attack \rangle$ be an argumentation framework. Consider the following preferential 3-labeling semantics for $\mathcal{AF}$:

- $\mathcal{CMP}(\mathcal{AF}) = \langle \mathsf{Lab}^3(Args), \mathsf{CMP}(\mathcal{AF}), \emptyset \rangle$.
- $\mathcal{GRND}(\mathcal{AF}) = \langle \mathsf{Lab}^3(Args), \mathsf{CMP}(\mathcal{AF}), <^{\mathsf{min}}_{\mathsf{in}} \rangle$,
  where $<^{\mathsf{min}}_{\mathsf{in}}$ is defined by $lab_1 <^{\mathsf{min}}_{\mathsf{in}} lab_2$ iff $\mathsf{In}(lab_1) \subsetneq \mathsf{In}(lab_2)$.

- $\mathcal{PREF}(\mathcal{AF}) = \langle \mathsf{Lab}^3(Args), \mathsf{CMP}(\mathcal{AF}), <^{\mathsf{max}}_{\mathsf{in}} \rangle$,
where $<^{\mathsf{max}}_{\mathsf{in}}$ is defined by $lab_1 <^{\mathsf{max}}_{\mathsf{in}} lab_2$ iff $\mathsf{In}(lab_1) \supsetneq \mathsf{In}(lab_2)$.
- $\mathcal{ST}(\mathcal{AF}) = \langle \mathsf{Lab}^3(Args), \mathsf{CMP}(\mathcal{AF}) \cup \mathsf{EM}(Args), \emptyset \rangle$,
where $\mathsf{EM}(Args) = \bigcup_{A \in Args} \{\mathsf{val}(A, \mathsf{in}) \vee \mathsf{val}(A, \mathsf{out})\}$. [3]
- $\mathcal{SST}(\mathcal{AF}) = \langle \mathsf{Lab}^3(Args), \mathsf{CMP}(\mathcal{AF}), <^{\mathsf{min}}_{\mathsf{undec}} \rangle$,
where $<^{\mathsf{min}}_{\mathsf{undec}}$ is defined by $lab_1 <^{\mathsf{min}}_{\mathsf{undec}} lab_2$ iff $\mathsf{Undec}(lab_1) \subsetneq \mathsf{Undec}(lab_2)$.

Proposition 14 *An argument $A \in Args$ skeptically follows from $\mathcal{AF}$ under complete (resp. grounded, preferred, stable, semi-stable) semantics iff* $\mathsf{val}(A, \mathsf{in})$ *is a skeptical conclusion of $\mathcal{AF}$ according to $\mathcal{CMP}(\mathcal{AF})$ (resp., according to $\mathcal{GRND}(\mathcal{AF})$, $\mathcal{PREF}(\mathcal{AF})$, $\mathcal{ST}(\mathcal{AF})$, $\mathcal{SST}(\mathcal{AF})$).*

Proof. We show for instance the case of skeptical acceptance under the grounded semantics; The other cases are proved similarly. It is easy to see that the models of $\mathsf{CMP}(\mathcal{AF})$ are the complete labelings of $\mathcal{AF}$. Thus, the valuations in $mod^3_{<^{\mathsf{min}}_{\mathsf{in}}}(\mathsf{CMP}(\mathcal{AF}))$ are exactly the grounded labelings of $\mathcal{AF}$. It follows that A is a skeptical conclusion of $\mathcal{AF}$ according to $\mathcal{GRND}(\mathcal{AF})$ iff all the grounded labelings of $\mathcal{AF}$ satisfy $\mathsf{val}(A, \mathsf{in})$, iff A is labeled in by every grounded labeling of $\mathcal{AF}$, iff A skeptically follows from $\mathcal{AF}$ according to the grounded semantics. □

Proposition 15 *An argument $A \in Args$ credulously follows from $\mathcal{AF}$ under complete (resp. grounded, preferred, stable, semi-stable) semantics iff* $\mathsf{val}(A, \mathsf{in})$ *is a credulous conclusion of $\mathcal{AF}$ according to $\mathcal{CMP}(\mathcal{AF})$ (resp., according to $\mathcal{GRND}(\mathcal{AF})$, $\mathcal{PREF}(\mathcal{AF})$, $\mathcal{ST}(\mathcal{AF})$, $\mathcal{SST}(\mathcal{AF})$).*

Proof. Similar to that of Proposition 14. □

Note 16 Other types of semantics, such as stage, ideal and eager semantics (see [5]), are also representable by 3-valued semantics in the form described in Definition 4. For these semantics one has to represent the condition $\mathsf{CMP}(\mathcal{AF})$ in terms of quantified Boolean formulas (that is, formulas involving only propositional languages and quantifications over propositional variables). We refer to [4] for the details.

3.2. Semantics of Constrained Argumentation Frameworks

In [11], Coste-Marquis, Devred and Marquis introduced the notion of a constrained argumentation frameworks (CAFs), with the aim of handling constraints on the evaluation of an argumentation framework. In this section we show that their approach fits together with our setting in a natural way.

Note 17 The formalism in [11] is expressed by sets of arguments (extensions) instead of labelings, and uses a similar language to ours , with the same interpretations for the connectives $\vee$ and $\neg$. To keep the presentation uniform and coherent, in what follows we adjust the notations and notions used in [11] to our setting.

[3] That is, $\mathsf{EM}(Args)$ 'excludes the middle label', undec.

Definition 18 [11] A *constrained argumentation framework* is a triple $\mathcal{CAF} = \langle Args, Attack, \mathcal{IC}\rangle$, where $\langle Args, Attack\rangle$ is an argumentation framework and $\mathcal{IC}$ is a set of formulas in $\mathcal{L}^3_{Args}$.

- lab is an *admissible labeling* of $\mathcal{CAF}$ if it is an admissible labeling of $\langle Args, Attack\rangle$ and an element in $mod^3_{Args}(\mathcal{IC})$.
- lab is a *preferred labeling* of $\mathcal{CAF}$ if it is an admissible labeling of $\mathcal{CAF}$ and $\mathsf{In}(lab)$ is maximal in $\{\mathsf{In}(lab) \mid lab \text{ is an admissible labelling of } \mathcal{CAF}\}$.
- lab is a *(semi-)stable labeling* of $\mathcal{CAF}$ if it is a (semi-)stable extension of $\langle Args, Attack\rangle$ and an element in $mod^3_{Args}(\mathcal{IC})$.

Definition 19 Let $\mathcal{CAF} = \langle Args, Attack, \mathcal{IC}\rangle$ and let $A \in Args$.

- A *skeptically follows* from $\mathcal{CAF}$ by admissible (resp. preferred, stable, semi-stable) semantics, if $lab(A) = \mathsf{in}$ for *every* admissible (resp. preferred, stable, semi-stable) 3-valued labeling lab of $\mathcal{CAF}$.
- A *credulously follows* from $\mathcal{CAF}$ by admissible (resp. preferred, stable, semi-stable) semantics, if $lab(A) = \mathsf{in}$ for *some* admissible (resp. preferred, stable, semi-stable) 3-valued labeling lab of $\mathcal{CAF}$.

Below we show that preferential 3-valued labeling semantics captures also the semantics of constrained argumentation frameworks.

Definition 20 Let $\mathcal{AF} = \langle Args, Attack\rangle$ and $\mathcal{CAF} = \langle Args, Attack, \mathcal{IC}\rangle$ be an argumentation framework and a constrained argumentation framework, respectively. Consider the following theory in the language $\mathcal{L}^3_{Args}$ for 3-valued argumentation:

$$\mathsf{ADM}(\mathcal{AF}) = \bigcup_{A \in Args} \left\{ \begin{array}{l} \mathsf{val}(A, \mathsf{in}) \supset \bigwedge_{B \in A^-} \mathsf{val}(B, \mathsf{out}), \\ \mathsf{val}(A, \mathsf{out}) \supset \bigvee_{B \in A^-} \mathsf{val}(B, \mathsf{in}), \\ \mathsf{val}(A, \mathsf{undec}) \supset \neg \bigvee_{B \in A^-} \mathsf{val}(B, \mathsf{in}) \end{array} \right\}.$$

Accordingly, we defined the following preferential 3-valued labeling semantics:

- $\mathcal{ADM}(\mathcal{CAF}) = \langle \mathsf{Lab}^3(Args), \mathsf{ADM}(\mathcal{AF}) \cup \mathcal{IC}, \emptyset\rangle$,
- $\mathcal{PREF}(\mathcal{CAF}) = \langle \mathsf{Lab}^3(Args), \mathsf{ADM}(\mathcal{AF}) \cup \mathcal{IC}, <^{\mathsf{max}}_{\mathsf{in}}\rangle$,
- $\mathcal{ST}(\mathcal{CAF}) = \langle \mathsf{Lab}^3(Args), \mathsf{CMP}(\mathcal{AF}) \cup \mathsf{EM}(Args) \cup \mathcal{IC}, \emptyset\rangle$.
- $\mathcal{SST}(\mathcal{CAF}) = \langle \mathsf{Lab}^3(Args), \mathsf{CMP}(\mathcal{AF}) \cup \mathcal{IC}, <^{\mathsf{min}}_{\mathsf{undec}}\rangle$.

Like the case of standard semantics, we have the following results (cf. Propositions 14 and 15).

Proposition 21 *$A \in Args$ skeptically follows from $\mathcal{CAF}$ under admissible (resp. preferred, stable, semi-stable) semantics iff* $\mathsf{val}(A, \mathsf{in})$ *is a skeptical conclusion of $\mathcal{CAF}$ according to $\mathcal{ADM}(\mathcal{CAF})$ (resp. according to $\mathcal{PREF}(\mathcal{CAF})$, $\mathcal{ST}(\mathcal{CAF})$, $\mathcal{SST}(\mathcal{CAF})$).*

Proposition 22 *$A \in Args$ credulously follows from $\mathcal{CAF}$ under admissible (resp. preferred, stable, semi-stable) semantics iff* $\mathsf{val}(A, \mathsf{in})$ *is a credulous conclusion of $\mathcal{CAF}$ according to $\mathcal{ADM}(\mathcal{CAF})$ (resp. according to $\mathcal{PREF}(\mathcal{CAF})$, $\mathcal{ST}(\mathcal{CAF})$, $\mathcal{SST}(\mathcal{CAF})$).*

3.3. Fallback Belief Semantics

It is well-known that every argumentation framework has a complete labeling (and so also a grounded and a preferred labeling). Such a labeling is a 3-valued valuation that satisfies the theory $\mathsf{CMP}(\mathcal{AF})$ in Definition 12. However, when integrity constraints are introduced the existence of complete labelings is not guaranteed anymore. As we saw in Section 3.2, in this case some approaches give up completeness and require only admissibility. A more radical step, taken e.g. in [8], is to give up admissibility and require only conflict-freeness. By Definition 6 these are the 3-valued labelings satisfying the following theory:

$$\mathsf{CF}(\mathcal{AF}) = \bigcup_{A \in Args} \left\{ \begin{array}{l} \mathsf{val}(A, \mathsf{in}) \supset \bigwedge_{B \in A^-} \mathsf{val}(B, \mathsf{out}), \\ \mathsf{val}(A, \mathsf{undec}) \supset \neg \bigvee_{B \in A^-} \mathsf{val}(B, \mathsf{in}) \end{array} \right\}.$$

The following relation is considered in [8] as a preference criterion among conflict-free labelings. Intuitively, it counts (and minimizes) the number of violations of the conditions for complete labelings that can be made by conflict-free labelings.

Definition 23 Let $\mathcal{AF} = \langle Args, Attack \rangle$ be an argumentation framework. Given a 3-valued conflict-free labeling lab for $\mathcal{AF}$, we denote by $|\mathsf{Illegal}(lab)|$ the size of the set $\mathsf{Illegal}(lab)$, considered in Definition 10. Accordingly, we define: $lab_1 <_{\mathsf{illeg}} lab_2$ iff $|\mathsf{Illegal}(lab_1)| < |\mathsf{Illegal}(lab_2)|$.

It follows that the fallback belief semantics considered in [8] can be described in terms of the preferential 3-labeling semantics

$$\mathcal{FBB}(\mathcal{AF}) = \langle \mathsf{Lab}^3(Args), \mathsf{CF}(\mathcal{AF}) \cup \mathcal{IC}(Args), <_{\mathsf{illeg}} \rangle,$$

where $\mathcal{IC}(Args)$ are the integrity constraints that the arguments should satisfy, expressed in the language $\mathcal{L}^3_{Args}$.

3.4. Conflict-Tolerant Semantics

Another type of semantics for abstract argumentation frameworks that can be represented by preferential n-labeling semantics is the conflict-tolerant approach introduced in [1,3]. This time, the underlying semantics is based on 4-valued labeling, where in addition to acceptance (in) and rejection (out), the two other labelings intuitively indicate lack of information (none) and contradictory information (both). Below are rationality postulates for 4-valued labelings (see [1,3]):

- **pIn** if $lab(A) = \mathsf{in}$ then for every $B \in A^-$ it holds that $lab(B) = \mathsf{out}$.
- **pOut** if $lab(A) = \mathsf{out}$ then there is $B \in A^-$ such that $lab(B) \in \{\mathsf{in}, \mathsf{both}\}$.
- **pBoth** if $lab(A) = \mathsf{both}$ then for every $B \in A^-$ $lab(B) \in \{\mathsf{out}, \mathsf{both}\}$ and there is $B \in A^-$ such that $lab(B) = \mathsf{both}$.
- **pNone** if $lab(A) = \mathsf{none}$ then for every $B \in A^-$, $lab(B) \in \{\mathsf{out}, \mathsf{none}\}$.

These conditions are expressible by the following theory in $\mathcal{L}^4_{Args}$:

$$\mathsf{CT}(\mathcal{AF}) = \bigcup_{A \in Args} \left\{ \begin{array}{l} \mathsf{val}(A,\mathsf{in}) \supset \bigwedge_{B \in A^-} \neg\mathsf{val}(B,\mathsf{out}), \\ \mathsf{val}(A,\mathsf{out}) \supset \big(\bigvee_{B \in A^-} \big(\mathsf{val}(B,\mathsf{in}) \vee \mathsf{val}(B,\mathsf{both})\big) \\ \mathsf{val}(A,\mathsf{both}) \supset \big(\bigwedge_{B \in A^-} \big(\mathsf{val}(B,\mathsf{out}) \vee \mathsf{val}(B,\mathsf{both})\big) \wedge \\ \qquad\qquad \bigvee_{B \in A^-} \mathsf{val}(B,\mathsf{both})\big) \\ \mathsf{val}(A,\mathsf{none}) \supset \big(\bigwedge_{B \in A^-} \big(\mathsf{val}(B,\mathsf{out}) \vee \mathsf{val}(B,\mathsf{none})\big) \end{array} \right\}$$

The labeling both implies that in some situations contradictory arguments must be accepted. Yet, while inconsistent belief about certain arguments is sometimes unavoidable, this is usually not desirable and should be avoided as much as possible. This is the intuition behind the following order relation: $lab_1 <^{\mathsf{min}}_{\mathsf{both}} lab_2$ iff $\mathsf{Both}(lab_1)| \subsetneq \mathsf{Both}(lab_2)$.[4]

It follows that the conflict-tolerant semantics described in [1,3] can be described in terms of the following preferential 4-labeling semantics:

$$\mathcal{CT}(\mathcal{AF}) = \langle \mathsf{Lab}^4(Args), \mathsf{CT}(\mathcal{AF}), <^{\mathsf{min}}_{\mathsf{both}} \rangle.$$

Note 24 In [2] consistency tolerance semantics is adjusted to handle integrity constraints. Just as we did in Section 3.2, it is possible to capture that semantics in our setting by augmenting $\mathcal{CT}(\mathcal{AF})$ with a set $\mathcal{IC}(Args)$ of integrity constraints that are specified in the language $\mathcal{L}^4_{Args}$. The preferential 4-valued labeling semantics that is obtained in this case is the following:

$$\mathcal{CT}(\mathcal{CAF}) = \langle \mathsf{Lab}^4(Args), \mathsf{CT}(\mathcal{AF}) \cup \mathcal{IC}(Args), <^{\mathsf{min}}_{\mathsf{both}} \rangle,$$

where $\mathcal{CAF} = \langle Args, Attack, \mathcal{IC} \rangle$ is the corresponding constraint argumentation framework.

3.5. Jakobovits and Vermeir's Labeling

In an attempt to satisfactorily handle arguments that, directly or indirectly, contradict themselves, Jakobovits and Vermeir [14] introduced another family of argumentation semantics that are based on four-valued labelings. The latter are complete functions on the set of arguments that satisfy the following conditions expressed by formulas in $\mathcal{L}^4_{Args}$:

$$\bigcup_{A \in Args} \left\{ \begin{array}{l} \big(\mathsf{val}(A,\mathsf{out}) \vee \mathsf{val}(A,\mathsf{none}\big) \supset \big(\bigvee_{B \in A^-} \big(\mathsf{val}(B,\mathsf{in}) \vee \mathsf{val}(B,\mathsf{both})\big) \\ \big(\mathsf{val}(A,\mathsf{in}) \vee \mathsf{val}(A,\mathsf{both}\big) \supset \big(\bigwedge_{B \in A^-} \big(\mathsf{val}(B,\mathsf{out}) \vee \mathsf{val}(B,\mathsf{none})\big) \\ \big(\mathsf{val}(A,\mathsf{in}) \vee \mathsf{val}(A,\mathsf{both}\big) \supset \big(\bigwedge_{B \in A^+} \big(\mathsf{val}(B,\mathsf{out}) \vee \mathsf{val}(B,\mathsf{none})\big) \end{array} \right\}.$$

Complete labelings are labelings that in addition satisfy the following condition:

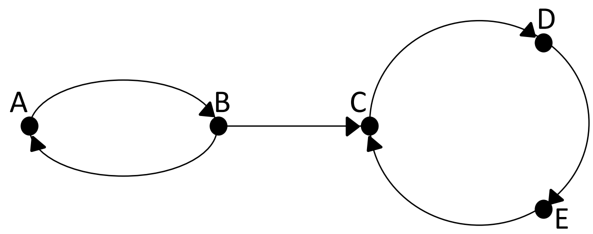
$$\bigcup_{A \in Args} \big(\mathsf{val}(A,\mathsf{in}) \vee \mathsf{val}(A,\mathsf{out}) \vee \mathsf{val}(A,\mathsf{both})\big).$$

Preference criteria are then developed for selecting the most 'robust' labelings among the complete ones (see [14] for the details).

[4] Recall that $\mathsf{Both}(lab) = \{A \in Args \mid lab(A) = \mathsf{both}\}$.

4. An Example

Let us illustrate the preferential n-labeling semantics considered here, using the argumentation framework $\mathcal{AF}$ in the figure below. As usual, the framework is represented by a directed graph in which the arguments are the nodes and attacks are the arrows.



- The $\mathcal{CMP}$-labelings of $\mathcal{AF}$ are the following:

	A	B	C	D	E
lab_1	undec	undec	undec	undec	undec
lab_2	in	out	undec	undec	undec
lab_3	out	in	out	in	out

- In the notation of the previous item, lab_1 is the $<_{\mathsf{in}}^{\mathsf{min}}$-preferred one, so it is the (only) $\mathcal{GRND}$-labeling of $\mathcal{AF}$, while both lab_2 and lab_3 are $<_{\mathsf{in}}^{\mathsf{max}}$-preferred, thus they are the $\mathcal{PREF}$-labelings of $\mathcal{AF}$. The latter are also the $\mathcal{ST}$-labelings and the $\mathcal{SST}$-labelings of $\mathcal{AF}$.

- Denote by $\mathcal{CAF}_1$ the constraint argumentation framework that consists of $\mathcal{AF}$ and the constraint $\mathcal{IC}_1 = \{\mathsf{val}(B, \mathsf{in})\}$. Since only lab_3 satisfies this constraint, it is the only $\mathcal{ST}$-labeling of $\mathcal{CAF}_1$.

- Let $\mathcal{CAF}_2$ be the constraint argumentation framework that consists of $\mathcal{AF}$ and the integrity constraints $\mathcal{IC}_2 = \{\mathsf{val}(B, \mathsf{in}), \mathsf{val}(E, \mathsf{in})\}$. This time there is no $\mathcal{CMP}$-labeling (nor $\mathcal{ST}$-labeling) for $\mathcal{CAF}_2$. Yet, using fallback belief semantics, we have that the (unique) $\mathcal{FBB}$-labeling of $\mathcal{CAF}_2$ is $lab(A) = lab(C) = lab(D) = \mathsf{out}$, $lab(B) = lab(E) = \mathsf{in}$. Note that only D is illegally labeled in lab. It can be checked that there is no other labeling that satisfies the constraints and has a set with a smaller or equal number of arguments that are illegally labeled.

- Since A and B attack each other, there are no conflict-free labelings for $\mathcal{AF}$ (nor for any constraint argumentation framework that is based on $\mathcal{AF}$) in which both of these arguments are mutually accepted. Conflict-tolerant semantics supports this situation. Let $\mathcal{CAF}_3$ be the constraint argumentation framework that consists of $\mathcal{AF}$ and the constraints $\mathcal{IC}_3 = \{\mathsf{acceptl}(A), \mathsf{acceptl}(B)\}$, where $\mathsf{accept}(\Psi) = \mathsf{val}(\Psi, \mathsf{in}) \vee \mathsf{val}(\Psi, \mathsf{both})$. Using conflict-tolerant semantics, we have that the (single) $\mathcal{CT}$-labeling of $\mathcal{CAF}_3$ is $lab(A) = lab(B) = \mathsf{both}$, $lab(C) = lab(E) = \mathsf{out}$, and $lab(D) = \mathsf{in}$.

5. Implementation by ASP

Preferential n-labeling semantics is not only representable by formal languages, but in many cases it is also programmable and computable. We demonstrate this by answer-set programming (ASP), which has been shown very useful for reasoning with different kinds of non-monotonic formalisms in general and argumentation semantics in particular (see, e.g., [10,13,16]).[5]

Generally, in order to reason with a preferential n-labeling semantics $\mathcal{SEM}(\mathcal{AF}) = \langle \mathsf{Lab}^n(Args), \mathsf{Cond}(\mathcal{AF}), < \rangle$ by answer-set programs, both the conditions in $\mathsf{Cond}(\mathcal{AF})$ and the preferential order $<$ must be computable. For this, the formers are usually represented by clause rules and the latter is expressed by a relation which is minimized. We refer, e.g., to [13,16] for some discussions on how this can be done in case of semantics like those considered in Section 3.1. ASP computations for the semantics discussed in Sections 3.3 and 3.4 are illustrated in the encodings shown in Figures 1 and 2. In these encodings we assume that the argumentation framework is represented by statements of the form `arg(X)` for every argument X and `att(X,Y)` for every attack. Furthermore, the constraints are assumed to be encoded by statements that assign labels to arguments.

Example 25 The answer-set program in Figure 1 computes fallback semantics.

```
<-- Framework encoding here -->
1 { in(X), out(X), undec(X) } 1 :- arg(X).
out(Y) :- att(X, Y), in(X).
out(X) :- att(X, Y), in(Y).
legally_out(X) :- out(X), att(Y, X), in(Y).
legally_undec(X) :- undec(X), att(Y, X), undec(Y).
illegally_out(X) :- out(X), not legally_out(X).
illegally_undec(X) :- undec(X), not legally_undec(X).
illegal(X) :- illegally_out(X).
illegal(X) :- illegally_undec(X).
<-- Constraint encoding here -->
# minimize { illegal(X) }.
```

Figure 1. Fallback semantics computation

Example 26 The program in Figure 2 computes conflict-tolerant semantics.

6. Conclusion

In this paper we introduced a general and uniform approach for representing a variety of labeling semantics for (constrained) argumentation frameworks, and hinted on their computation by answer-set programs. Future work involves a development of automated tools for supporting a more comprehensive implementation of our setting and a comparative study on the expressive power of different preferential n-labeling semantics, as well as their suitability for properly handling practical problems.

[5] Other tools for computerized reasoning may be incorporated as well. See, e.g., [4,6].

```
<-- Framework encoding here -->
1  in(X), out(X), both(X), none(X)  1 :- arg(X).
:- in(X), not out(Y), att(Y,X).
:- out(X), not in(Y), not both(Y), att(Y, X).
:- out(X), not in(Y), not none(Y), att(Y, X).
:- both(X), in(Y), att(Y, X).
:- both(X), none(Y), att(Y, X).
:- both(X), not both(Y), att(Y, X).
:- none(X), in(Y), att(Y, X).
 :- none(X), both(Y), att(Y, X).
 <-- Constraint encoding here -->
 #  minimize [ both(X) ].
```

Figure 2. Conflict-tolerant semantics computation

References

[1] O. Arieli. Conflict-tolerant semantics for argumentation frameworks. In *Proc. JELIA'12*, LNCS 7519, pages 28–40. Springer, 2012.

[2] O. Arieli. Towards constraints handling by conflict tolerance in abstract argumentation frameworks. In *Proc. FLAIRS'2013*, pages 585–590. AAAI Press, 2013.

[3] O. Arieli. On the acceptance of loops in argumentation frameworks. *Journal of Logic and Computation*, 2014. doi:10.1093/logcom/exu009.

[4] O. Arieli and M. W. A. Caminada. A QBF-based formalization of abstract argumentation semantics. *Journal of Applied Logic*, 11(2):229–252, 2013.

[5] P. Baroni, M. W. A. Caminada, and M. Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(4):365–410, 2011.

[6] Ph Besnard, A. Hunter, and S. Woltran. Encoding deductive argumentation in quantified boolean formulae. *Artificial Intelligence.*, 173(15):1406–1423, 2009.

[7] R. Booth, S. Kaci, T. Rienstra, and L. van der Torre. Conditional acceptance functions. In *Proc. COMMA'12*, volume 245 of *Frontiers in Artificial Intelligence and Applications*, pages 470–477. IOS Press, 2012.

[8] R. Booth, S. Kaci, T. Rienstra, and L. van der Torre. A logical theory about dynamics in abstract argumentation. In *Proc. SUM'13*, LNCS 8078, pages 148–161. Springer, 2013.

[9] M. W. A. Caminada and D. M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009. Special issue: New ideas in argumentation theory.

[10] G. Charwat, J. P. Wallner, and S. Woltran. Utilizing ASP for generating and visualizing argumentation frameworks. *CoRR*, abs/1301.1388, 2013.

[11] S. Coste-Marquis, C. Devred, and P. Marquis. Constrained argumentation frameworks. In *Proc. KR'06*, pages 112–122. AAAI Press, 2006.

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

[13] U. Egly, S. A. Gaggl, and S. Woltran. Answer-set programming encodings for argumentation frameworks. *Argument and Computation*, 1(2):144–177, 2010.

[14] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 9(2):215–261, 1999.

[15] I. Rahwan and G. R. Simari. *Argumentation in Artificial Intelligence*. Springer, 2009.

[16] F. Toni and M. Sergot. Argumentation and answer set programming. In M. Balduccini and T. Son, editors, *Logic Programming, Knowledge Representation, and Nonmonotonic Reasoning*, LNCS 6565, pages 164–180. Springer, 2011.

[17] B. Verheij. Two approaches to dialectical argumentation: admissible sets and argumentation stages. In *Proc. NAIC'96*, pages 357–368, 1996.