

United We Stand: Accruals in Strength-based Argumentation

Julien Rossit

LIPADE, Université de Paris, France
julien.rossit@u-paris.fr

Jean-Guy Mailly

LIPADE, Université de Paris, France
jean-guy.mailly@u-paris.fr

Yannis Dimopoulos

Department of Computer Science, University of Cyprus, Cyprus
yannis@cs.ucy.ac.cy

Pavlos Moraitis

LIPADE, Université de Paris, France
Argument Theory, France
pavlos.moraitis@u-paris.fr

Abstract

Argumentation has been an important topic in knowledge representation, reasoning and multi-agent systems during the last twenty years. In this paper, we propose a new abstract framework where arguments are associated with a strength, namely a quantitative information which is used to determine whether an attack between arguments succeeds or not. Our *Strength-based Argumentation Framework* (StrAF) combines ideas of Preference-based and Weighted Argumentation Frameworks in an original way, which permits to define acceptability semantics sensitive to the existence of accruals between arguments. The question of accruals arises in situations where several arguments defending the same position (but from different points of view) against another argument are unable to individually defeat this argument, but could do it collectively if they combine their strengths. We investigate some of the theoretical and computational properties of our new framework and semantics, and present a reasoning algorithm that is based on a translation of the problem into pseudo-boolean constraint satisfaction. This paper proposes an intuitive framework which allows strength compensations in an argumentation context where attacks may not succeed, completed by an approach which detects accruals throughout the reasoning process without requiring the elicitation of all compensatory combinations of arguments as an input.

2012 ACM Subject Classification Computing methodologies → Nonmonotonic, default reasoning and belief revision

Keywords and phrases Abstract argumentation, Argument strength, Accruals, Collective defeat

Category Technical Report

Note: This paper has been accepted for publication in *Argument & Computation*.

Introduction

Argumentation has been a main research topic in artificial intelligence for the last twenty years, with applications in various domains such as decision making [4], automated negotiation [33], reasoning with inconsistent knowledge [13], legal reasoning [11], and multi-agent systems [54]. A lot of works have been proposed, mainly based on the influential argumentation framework (AF) of Dung [35]. An AF is characterized by a set of abstract entities called *arguments* and an *attack relation* between arguments. A series of semantics have been defined to determine which arguments are acceptable, usually by computing sets of arguments called *extensions* that represent arguments that are compatible with each other [7]. Besides situations where

arguments are reasons for believing some claims, AFs allow to model different situations where conflicts arise between pieces of information, and to obtain some valuable conclusions.

Reasoning with such an AF supposes that there is no ambiguity about the nature of conflicts and how to resolve them. However, and even when the nature of conflicts is certain, different agents may apply contrasting policies to resolve these conflicts, depending on the relative priorities they associate with arguments. These relative priorities may be represented as a relation of preference between arguments [3], or by a relation of preference between the values attached to arguments [10]. In these works, the reasoning process is achieved in two steps: first a *defeat* relation is defined as the combination of the initial attack relation and the preference relation. Then, an extension-based semantics of Dung is applied on the graph obtained from the set of arguments combined with the resulting defeat relation. However, none of these frameworks allows to finely compare arguments with respect to a quantitative strength. Intuitively speaking, considering a strength-based framework allows to induce priorities among arguments by merely associating a weight to each argument, and specially does not require to examine each possible pairs of arguments.

Assigning a quantitative information to arguments has been considered in different contexts. First, [15] associates arguments with a numerical value that represents "fuzziness, probability or a preference in general"; the extensions given by Dung's semantics are then refined depending on the weights of the arguments that belong to them. Then, a different approach considers arguments mapped to a quantitative strength in another context: instead of defining arguments status through the notion of extension, rankings or graduations of arguments can be defined by analyzing the arguments' intrinsic strength and the attack relation between arguments [2]. Contrary to preference-based argumentation, these weighted argumentation frameworks defined in [15, 2] do not consider a notion of defeat in the reasoning process. Moreover, none of the options presented above permits to consider collective attacks [57], and especially *accrual* of arguments.

The idea of accruals, *i.e.* arguments that cannot on their own defeat their target but together can succeed, has been initially studied in the literature of philosophy. While, at first, some doubts have been expressed about the meaning of arguments accrual [60, 59], more recently assumptions under which accruals of arguments make sense have been pointed out [56]. However the question of whether arguments should accrue or not is out of the scope of this paper. Indeed, this discussion may be decisive when considering arguments as reasons for supporting a claim, but we do not restrict our study to such applications and opt instead for a general view of argumentation as a formal tool to deal with conflicting pieces of information. In addition and more specifically, the question of accruals of reasons has been addressed in the AI literature (see [66, 62, 57] for examples). Although these studies differ from this paper since they settle in a rule-based argumentation context, corresponding approaches furthermore require to express explicitly *a priori* all sets of accruals to take them into account throughout the reasoning process. To the best of our knowledge, the question of how accruals can be detected during the reasoning process has not been addressed in the literature.

The aim of this paper is to study this question. More precisely, we propose an approach which:

- allows strength compensations in an argumentation context where attacks may not succeed;
- detects accruals of arguments during the reasoning process without requiring their explicit elicitation in the model as an input.

To this purpose, we define the *Strength-based Argumentation Framework* which combines in

an original way the quantitative strength expressed in weighted argumentation frameworks on one hand, and the notion of defeat relation introduced in the contexts of preference-based and value-based argumentation on another hand. Then, we generalize the notion of defeat to define a notion of *collective defeat*. In the following, we use the term of *accrual* to identify a set of arguments that attack a same target. While arguments in this set might not be able to individually defeat their common target, they could nevertheless achieve the defeat by combining their strength. The representation proposed in this paper allows to compute the strength of an existing accrual, and consequently to decide about the outcome of a conflict between an accrual and an individual argument (or between two accruals) by applying a slightly adapted version of the Dung acceptability semantics.

The paper is organized as follows. Section 1 recalls the basic notions of abstract argumentation. In Section 2, we propose an intuitive framework for representing strength of arguments, and we define formally a defeat-based semantics for our Strength-based Argumentation Frameworks (StrAFs). Then Section 3 introduces accrual of arguments, and provides some theoretical properties of these new semantics that are sensitive to this notion of accrual and produce extensions that are usually ignored by existing semantics. In Section 4, we study the complexity of the reasoning processes for first some particular sub-classes of StrAFs, and then for the general case. We show that, despite this reasoning mechanism may appear to be theoretically intractable, we can rely on the power of Pseudo-Boolean constraints solvers to compute extensions produced by our framework. Section 5 discusses in details existing related work. Finally, Section 6 concludes the paper and describes some tracks for future research.

1 Preliminaries

Most of the studies addressed to the literature of argumentation for the last twenty years have been inspired by the work of [35]. This section presents the basics of abstract argumentation frameworks.

An *argumentation framework* (AF), as introduced by Dung in [35], is a pair $\langle \mathcal{A}, \mathcal{R} \rangle$, where \mathcal{A} is a set of *arguments*, and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an *attack relation*. The relation a attacks b , or b is attacked by a , is denoted by $(a, b) \in \mathcal{R}$. We focus on finite argumentation frameworks, *i.e.* AFs such that \mathcal{A} is a finite set.

In [35], different acceptability semantics have been introduced. These are based on two basic concepts, *defence* and *conflict-freeness*, defined as follows:

► **Definition 1.** (*Defence/Conflict-freeness*) Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation system. Let $S \subseteq \mathcal{A}$.

- S is conflict-free if and only if $\nexists a, b \in S$ s.t. $(a, b) \in \mathcal{R}$.
 - S defends $a \in \mathcal{A}$ if and only if $\forall b \in \mathcal{A}$, if $(b, a) \in \mathcal{R}$, then $\exists c \in S$ such that $(c, b) \in \mathcal{R}$.
- The set of all conflict-free sets of AF is denoted $\text{cf}(AF)$.

The basic idea behind these semantics is the following: for a rational agent, an argument a is acceptable if he can defend a against all attacks. All the arguments acceptable for a rational agent will be gathered in a so-called *extension*. An extension must satisfy a consistency requirement and must defend all its elements.

► **Definition 2.** (*Acceptability semantics*) Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be an AF and $S \in \text{cf}(AF)$.

- S is an admissible set if and only if S defends any element in S .
- S is a preferred extension if and only if S is a \subseteq -maximal admissible set.

σ	$\sigma(AF_{john})$
cf, ad	$\emptyset, \{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{F\}, \{A, C\}, \{A, D\}, \{A, E\}, \{A, F\},$ $\{B, C\}, \{B, D\}, \{B, E\}, \{B, F\}, \{C, E\}, \{C, F\}, \{D, E\}, \{D, F\}$ $\{A, C, E\}, \{A, C, F\}, \{A, D, E\}, \{A, D, F\},$ $\{B, C, E\}, \{B, C, F\}, \{B, D, E\}, \{B, D, F\}$
pr, st	$\{A, C, E\}, \{A, C, F\}, \{A, D, E\}, \{A, D, F\},$ $\{B, C, E\}, \{B, C, F\}, \{B, D, E\}, \{B, D, F\}$
σ	$\sigma(AF_{john}^2)$
cf, ad	$\emptyset, \{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{F\}, \{A, E\}, \{B, E\}, \{C, E\}, \{D, E\},$ $\{A, F\}, \{B, F\}, \{C, F\}, \{D, F\}, \{A, C\}, \{A, D\}, \{B, D\}, \{C, D\},$ $\{A, C, E\}, \{A, D, E\}, \{B, D, E\}, \{C, D, E\}, \{A, C, F\}, \{A, D, F\},$ $\{B, D, F\}, \{C, D, F\}, \{A, C, D\}, \{A, C, D, E\}, \{A, C, D, F\}$
pr, st	$\{A, C, D, E\}, \{A, C, D, F\}, \{B, D, E\}, \{B, D, F\}$

■ **Table 1** Extensions of AF_{john} and AF_{john}^2

to an intrinsic strength, that represents votes given by users [48], a certainty degree of arguments premises [12] or trustworthiness of the source of information [32]. A similar intuitive explanation is given in [55, 61] for weights associated with arguments, and in [37] for weights associated with attacks.

Game theory techniques have also been borrowed to determine what is the strength of an argument depending on the structure of the argumentation graph [53]. While this approach gives a concrete meaning to the strength of arguments, here the strength is the output of the process, in the spirit of gradual semantics [24, 8]. On the contrary, we want to have the strength as input of our reasoning process.

Concrete applications of argumentation frameworks with strength of arguments have been studied recently. Quantitative Argumentation Debates (QuAD frameworks) [9] associate arguments to a numerical strength (called base score), and their gradual semantics define a degree of acceptance for each argument. A variant of QuAD frameworks [64] has been used to give a novel method for opinion polling, with arguments base scores corresponding to users votes.

Quantitative Bipolar Argumentation Frameworks (QBAFs) [8] are a general abstract argumentation framework where arguments are related by both attacks and supports [25], and they are attached with a base score. Again, the semantics of these frameworks yields a degree of acceptance for each argument. In [27], QBAFs are used to aggregate reviews of movies from Internet database, in particular the notes given on the famous website RottenTomatoes are used as the base score of arguments.

An argumentation framework for persuasion is defined in [65], where arguments are associated with integer weights. However, these weights can be interpreted as the arguments weakness instead of the arguments strength, since they represent the cost of the action supported by the argument. Thus, the higher is the weight, the weaker is the argument.

In our running example, we show how strengths of arguments can be used to represent the utility for the agent that some arguments belong to the extensions (for instance, the money earned when performing some task). For a matter of simplicity, we will use natural numbers for modelling the strength of arguments in all our definitions and examples, although our approach can be extended with positive real numbers. However, the behaviour of our accrual-sensitive semantics may not be preserved in some specific cases, that are discussed in Section 3.1.

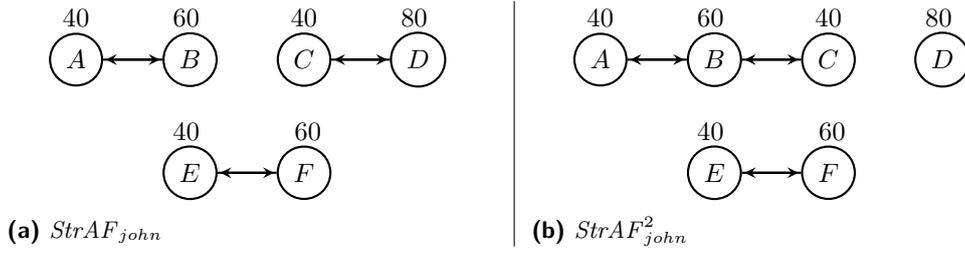
2.2 Formal Definition of StrAFs

Let us now formally introduce the Strength-based Argumentation Framework. Intuitively speaking, a StrAF is an argumentation framework where each argument is associated with a weight (here a non-negative integer) which represents its strength:

► **Definition 4.** (*Strength-based Argumentation Framework*) A Strength-based Argumentation Framework (*StrAF*) is a triple $\langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ where \mathcal{A} and \mathcal{R} are respectively arguments and attacks, and $\mathcal{S} : \mathcal{A} \rightarrow \mathbb{N}$ is a strength function.

In our framework, the higher is the weight, the stronger is the argument this weight is associated with. Let us illustrate this representation with the following example:

► **Example 5.** Let us continue Example 3. Depending of the size of the garden, John's salary for these works is not the same: houses A , B , C , D , E and F are worth respectively 40 Euros, 60 Euros, 40 Euros, 80 Euros, 40 Euros and 60 Euros. For each argument, the salary that John gets corresponds to the strength of the argument. So, from AF_{john} and AF_{john}^2 , we can define StrAFs $StrAF_{john} = \langle \mathcal{A}_{john}, \mathcal{R}_{john}, \mathcal{S}_{john} \rangle$ and $StrAF_{john}^2 = \langle \mathcal{A}_{john}, \mathcal{R}_{john}^2, \mathcal{S}_{john} \rangle$ described at Figure 2, where $\mathcal{S}(A) = \mathcal{S}(C) = \mathcal{S}(E) = 40$, $\mathcal{S}(B) = \mathcal{S}(F) = 60$ and $\mathcal{S}(D) = 80$. This means that an argument X is stronger than an argument Y if John earns more money when maintaining the garden of house X rather than house Y .



■ **Figure 2** Both Versions of John's Garden Situation

Now, we adapt Dung-style semantics to StrAFs. To do so, we borrow the notion of defeat relation from Preference-based AFs and Value-based AFs [3, 10].

► **Definition 6.** (*Defeat Relation and Extension Semantics*) Given $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ a StrAF, the defeat relation *Def* is defined by $Def = \{(a, b) \in \mathcal{R} \mid \mathcal{S}(b) \not\geq \mathcal{S}(a)\}$.

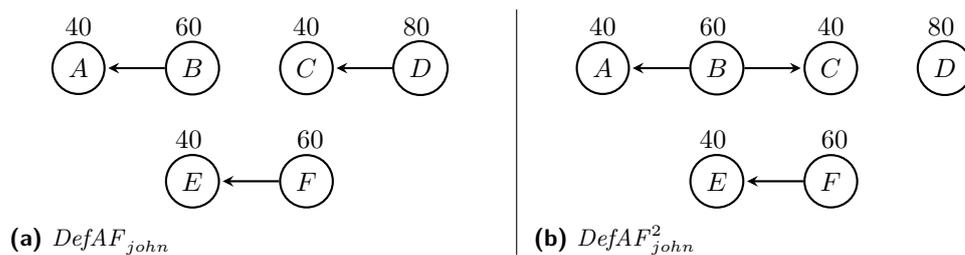
Extension-based semantics of StrAF can be defined by considering the extensions of the classical AF $DefAF = \langle \mathcal{A}, Def \rangle$, i.e. $\sigma(StrAF) = \sigma(DefAF)$ for $\sigma \in \{cf, ad, pr, st\}$.¹

This definition mimics the definition of Preference-based AFs, except that here the notion of preference is expressed through a quantitative measure. A similar notion can be found in [16] where the concept of defence is refined by taking advantage of weights associated with attack relations.

► **Example 7.** We continue Example 5. For both $StrAF_{john}$ and $StrAF_{john}^2$, we describe below the graphs corresponding to the defeat relation.

In the case of $DefAF_{john}$ (Figure 3a), we observe that using the defeat relation leads to an intuitive result: for any semantics $\sigma \in \{ad, pr, st\}$ as introduced in Definition 2, the single

¹ Of course, the same mechanism can be applied to other semantics that are not considered in this paper.



■ **Figure 3** Defeat Graphs for John's Garden Situation

extension is $\{B, D, F\}$, which is the best option for John (since he will earn 200 Euros). However, in the second situation depicted at Figure 3b, any semantics $\sigma \in \{\text{ad}, \text{pr}, \text{st}\}$ gives $\sigma(AF_{john}^2) = \{B, D, F\}$; this option is worth 200 Euros and is not optimal (indeed $\{A, C, D, F\}$ is an extension of AF_{john}^2 , and is worth 220 Euros).

These defeat-based semantics are relevant when the strength associated with an argument is individual and independent from other arguments, likewise the behavior of Preference-based AFs. In our case, the semantics may produce an extension which is not the most desirable outcome, like it is the case with $DefAF_{john}^2$. This is why we propose to define semantics sensitive to the notion of accrual of arguments.

3 Accrual of Arguments

Recent works have tackled the question of "quality versus quantity": is it worse for an argument to be attacked by only one strong attacker, or to be attacked by plenty of quite weak attackers? In the context of gradual semantics, this question is materialized by the principles of Quality Precedence and Cardinality Precedence. A Compensation principle is also stated to describe situations where several weak arguments have the same effect on their target as fewer strong arguments [1, 20, 2]. Although the context is not the same (since we consider extension-based semantics), a similar intuition leads to our definition of accruals: several arguments may be individually too weak to defeat their target, but their collective attack may be strong enough to compensate or even exceed the target's strength.

3.1 StrAFs with Accrual

In this paper, we assume that weights associated with arguments are commensurable. Roughly speaking, this means that these weights are comparable from an argument to another, and aggregating them (with *e.g.* a Sum-based operation) makes sense. This kind of assumptions suits well situations where, for instance, weights associated with arguments are provided by an expert or a group of experts sharing a same representation scale, or are regarded as rewards granted for the acceptance of some arguments. An example of the latter is provided by our running example.

Let us first formally define the notion of accrual of arguments:

► **Definition 8.** Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a StrAF. A set of arguments $\kappa \subseteq \mathcal{A}$ is called accrual if and only if $\exists c \in \mathcal{A}$ such that $\forall a \in \kappa, (a, c) \in \mathcal{R}$. Moreover, we say that κ is an accrual that attacks argument c .

Intuitively, an accrual is a set of arguments such that there exists an argument which is attacked by all arguments in this set. Let us illustrate this notion with our running example:

► **Example 9.** In $StrAF_{john}^2$ describing John's situation (see Example 5), there is (for instance) an accrual $\kappa = \{A, C\}$ attacking the argument B .

In the following, we choose to opt for a pessimistic view of attacking an accrual. An accrual is said to be attacked by an argument if and only if at least one of its arguments is attacked. Then, several definitions of *an accrual attacking another accrual* are possible. In what follows, we choose to focus on the following one, which corresponds again to the pessimistic case: an accrual κ attacks an accrual κ' if and only if there exists an argument in κ' such that all arguments of κ attack this argument. Formally:

► **Definition 10.** Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a $StrAF$, and $\kappa \subseteq \mathcal{A}$, $\kappa' \subseteq \mathcal{A}$ two accruals. Then, κ attacks κ' if and only if $\exists a \in \kappa'$ such that κ attacks a .

We then define the *collective strength* associated with an accrual κ , denoted by $\text{coval}(\kappa)$, as the combination of values associated with arguments. Formally :

► **Definition 11.** Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a $StrAF$ and $\kappa = \{a_1, \dots, a_n\} \subseteq \mathcal{A}$ be an accrual. Then the *collective strength* associated with κ is:

$$\text{coval}(\kappa) = \text{coval}(\mathcal{S}(a_1), \dots, \mathcal{S}(a_n))$$

where coval is an aggregation operator, i.e. a mapping from a vector of integers to an integer that satisfies:

(non-decreasingness) if $x_i \geq x'_i$, then $\text{coval}(x_1, \dots, x_i, \dots, x_n) \geq \text{coval}(x_1, \dots, x'_i, \dots, x_n)$;

(minimality) $\text{coval}(x_1, \dots, x_n) = 0$ if and only if $x_1 = \dots = x_n = 0$;

(identity) $\text{coval}(x) = x$;

(accrual) $\text{coval}(x_1, \dots, x_n) \leq \text{coval}(x_1, \dots, x_n, y)$.

These properties are quite natural for defining the strength of an accrual. Non-decreasingness means that, if $a \notin \kappa$ and $b \notin \kappa$, then the accrual $\kappa \cup \{a\}$ is stronger than the accrual $\kappa \cup \{b\}$ when a is stronger than b . Minimality states that the accumulation of arguments "without" strength (i.e. $\mathcal{S}(a_i) = 0$) does not create a collective strength from the void, and that non-null arguments cannot cancel each other's strength when they are taken together. Identity ensures that the "collective" strength of a singleton is exactly the intrinsic strength of the argument in this singleton. These properties are used for defining aggregation operators with other purposes, like defining extensions in AFs with weighted attacks [30]; they correspond to the basic properties of semirings [14, 16]. Finally, we consider the Accrual property: any set of arguments is at least as strong as all its subsets.

The coval operator may be a classical aggregation function like the sum \sum , the maximum \max or the weighted sum. Conversely, the product Π does not satisfy the minimality property. Furthermore, one can notice that if we extend our approach to positive real numbers for representing the strength, then the product Π operator does not satisfy the accrual properties if some values belong to the interval $[0, 1]$. Nevertheless, all the properties remain satisfied by Π if only natural or real numbers greater or equal to 1 are allowed.

To accommodate the notion of accrual we extend the semantics of $StrAF$ s to take collective defeat as follows.

We say that an argument a is collectively defeated by an accrual κ if and only if the collective strength associated with κ is greater or equal to the value associated with a .

► **Definition 12.** Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a $StrAF$, $a \in \mathcal{A}$, and coval an aggregation operator. Then, an accrual κ defeats a with respect to coval , denoted by $\kappa \triangleright_{\text{coval}} a$, if and only if

- $\kappa \subseteq \mathcal{A}$ is an accrual that attacks a ;
- $\text{coval}(\kappa) \geq \mathcal{S}(a)$.

If coval is clear from the context, we use $\kappa \triangleright a$ instead of $\kappa \triangleright_{\text{coval}} a$.

We can introduce the notion of an accrual that defeats another accrual as follows:

► **Definition 13.** Let $\text{StrAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a StrAF, coval an aggregation operator and $\kappa \subseteq \mathcal{A}$, $\kappa' \subseteq \mathcal{A}$ two accruals. Then κ defeats κ' , denoted by $\kappa \triangleright \kappa'$, if and only if there exists $a \in \kappa'$ such that $\kappa \triangleright a$.

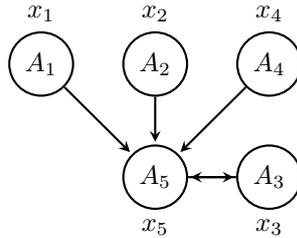
Roughly speaking, an accrual defeats another accrual if the first accrual induces a defeat against at least one argument of the second accrual.

In [16], the strength of an attack from a set S to a set S' of arguments is defined as the aggregation of strengths associated with all the attacks from any argument in S to any argument in S' . While this may look similar to defeats between accruals, these are actually different concepts. Indeed, since [16] do not associate strength with arguments, but with attacks, there is no comparison between strength associated respectively with S and S' . Moreover, in our approach, defeating any argument of an accrual is enough to consider the whole accrual as defeated. Nevertheless, one can notice that a subset of this defeated accrual can still be an undefeated accrual.

Let us illustrate the notion of collective attacks with our running example.

► **Example 14.** We continue Example 5. The natural aggregation operator here is $\text{coval} = \sum$, since the strength of an accrual is the money earned by John for maintaining several gardens (corresponding to the arguments in the accrual). We observe that, in $\text{StrAF}_{\text{john}}^2$, the accruals $\kappa_1 = \{B\}$ and $\kappa_2 = \{A, C\}$ defeat each other. Indeed, $\text{coval}(\kappa_1) = \mathcal{S}(B) = 60 \geq \mathcal{S}(A)$, while $\text{coval}(\kappa_2) = \mathcal{S}(A) + \mathcal{S}(C) = 80 \geq \mathcal{S}(B)$.

The following example illustrates our approach in a non-symmetrical abstract argumentation context.



■ **Figure 4** The StrAF that describes the situation of Joe and Jack

► **Example 15.** Let us consider the example provided in [57], composed of the following abstract arguments:

- A_1 : "Joe does not like Jack";
- A_2 : "There is a nail in Jack's antique coffee table";
- A_3 : "Joe hammered a nail into Jack's antique coffee table";
- A_4 : "Joe plays golf, so Joe has full use of his arms";
- A_5 : "Joe has no arms, so Joe cannot use a hammer, so Joe did not hammer a nail into Jack's antique coffee table".

As examples of attacks, one can on a first hand notice that the argument A_5 directly attacks (and thus defeats) the argument A_3 , whereas arguments A_3 and A_4 directly attack (and also defeat) the argument A_5 .

On another hand, authors in [57] discuss that arguments A_1 and A_2 attack the argument A_5 since they can provide the conclusion that Joe has hammered a nail into Jack's antique coffee table, but cannot separately defeat it. However, when considered together, A_1 and A_2 may constitute a stronger argument that defeats A_5 . These examples of attacks are illustrated by Figure 4.

In our framework, let x_i be the strength associated with the argument A_i , for $i \in \{1, \dots, 5\}$. Let us assume that $x_3 = x_5$ (since A_3 and A_5 defeat each other), and $x_4 \geq x_5$ (since A_4 defeat A_5). Several cases are then possible:

- either strengths associated with arguments are such that either $x_1 \geq x_5$ or $x_2 \geq x_5$: in this case, either the argument A_1 or the argument A_2 is considered solid enough to individually defeat the argument A_5 ;
- either strengths associated with arguments are such that $x_1 < x_5$, $x_2 < x_5$ and $x_1 + x_2 \geq x_5$: in this case arguments A_1 and A_2 are too weak to defeat individually argument A_5 but form an accrual to collectively defeat it;
- or finally strengths associated with arguments are such that $x_1 < x_5$, $x_2 < x_5$ and $x_1 + x_2 < x_5$: in this case arguments A_1 and A_2 cannot defeat A_5 neither individually nor collectively.

Roughly speaking, according to the importance one chooses to attach to these arguments and thus depending on the strengths associated with arguments A_1 , A_2 and A_5 , A_1 and A_2 can individually or collectively defeat A_5 or not.

3.2 Extension-based Semantics for StrAFS

In [35], different acceptability semantics have been introduced for computing the status of arguments. These are based on two basic concepts, *defence* and *conflict-freeness*, which can be adapted to the context of accruals.

The notion of conflict-freeness can be defined into two different senses, the *strong* and the *weak* conflict-freeness.

► **Definition 16.** (*Defence/Conflict-freeness*) Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a StrAF, coval an aggregation operator, and $S \subseteq \mathcal{A}$.

- S is strongly conflict-free if and only if $\nexists a, b \in S$ such that $(a, b) \in \mathcal{R}$.
- S is weakly conflict-free if and only if there are no accruals $\kappa_1 \subseteq S$ and $\kappa_2 \subseteq S$ such that $\kappa_1 \triangleright_{\text{coval}} \kappa_2$.
- S defends an element $a \in \mathcal{A}$ if and only if for all accruals $\kappa_1 \subseteq \mathcal{A}$, if $\kappa_1 \triangleright_{\text{coval}} a$, then there exists an accrual $\kappa_2 \subseteq S$ such that $\kappa_2 \triangleright_{\text{coval}} \kappa_1$.

Depending on the notion of conflict-freeness that is applied, two versions of acceptability semantics are derived, that are defined formally below.

► **Definition 17.** (*Acceptability semantics*) Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a StrAF, coval an aggregation operator, and S a strong (respectively weak) conflict free set of arguments.

- S is a strong (respectively weak) admissible set if and only if S defends all elements of S .
- S is a strong (respectively weak) preferred extension if and only if S is a \subseteq -maximal strong (respectively weak) admissible set.

- S is a strong (respectively weak) stable extension if and only if for each argument $a \in \mathcal{A} \setminus S$, there is an accrual $\kappa \subseteq S$ such that $\kappa \triangleright a$.

For any semantics $\sigma \in \{\text{cf}, \text{ad}, \text{pr}, \text{st}\}$, the set of strong (respectively weak) σ extension of $\text{StrAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ with respect to coval is denoted by $\sigma_S^{\text{coval}}(\text{StrAF})$ (respectively $\sigma_W^{\text{coval}}(\text{StrAF})$). We drop coval from the notations when it is clear from the context.

It is easy to see that a strong extension is a weak extension as well, since the absence of attacks in a set S (*i.e.* strong conflict-freeness) straightforwardly implies the absence of accrual in S defeating elements of S (*i.e.* weak conflict-freeness). Therefore $\sigma_S(\text{StrAF}) \subseteq \sigma_W(\text{StrAF})$ for any semantics σ .

We also remark a relation between weak semantics of StrAFs and semantics of AFs with collective attacks [57]. Indeed, if $\kappa \triangleright a$ holds in a StrAF, we can define an AF with collective attacks where the set of arguments κ attacks a (according to the notion of attacks defined in [57]). Thus, weak conflict-freeness (and the weak semantics based on weak conflict-freeness) coincide with the conflict-freeness and semantics for AFs with collective attacks. However, we notice that StrAFs allow a more modular representation of argument accrual, since they are only based on individual attacks and arguments strengths, and they do not need to be explicitly defined *a priori*.

We show that the usual relationship between semantics also holds for StrAFs.

► **Proposition 18.** (*Semantics Inclusion*) Let $\text{StrAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a StrAF and coval an aggregation operator. For $X \in \{S, W\}$, $\text{st}_X(\text{StrAF}) \subseteq \text{pr}_X(\text{StrAF}) \subseteq \text{ad}_X(\text{StrAF})$.

Proof. The fact that every strong (respectively weak) preferred extension is a strong (respectively weak) admissible extension is straightforward from the definition.

Let $E \subseteq \mathcal{A}$ be a strong (respectively weak) stable extension of $\text{StrAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ with respect to coval, *i.e.* $\forall a \in \mathcal{A} \setminus S$, there is an accrual $\kappa \subseteq E$ such that $\kappa \triangleright a$.

First we prove that E is a strong (respectively weak) admissible extension. Let $\kappa \subseteq \mathcal{A}$ be an accrual such that $\kappa \triangleright a$, with $a \in E$. There are two cases.

Case 1 $\kappa \subseteq E$ implies that E is neither strongly nor weakly conflict-free, this is a contradiction.

Case 2 $\kappa \not\subseteq E$ implies that $\exists b \in \kappa \setminus E$, and since E is a strong (respectively weak) stable extension, $E \triangleright b$ holds. Thus $E \triangleright \kappa$, and we conclude that E is a strong (respectively weak) admissible extension.

Now we prove that E is a strong (respectively weak) preferred extension. Using *reductio ad absurdum*, we suppose that E is not a strong (respectively weak) preferred extension, this means that $\exists E' \subseteq \mathcal{A}$ a strong (respectively weak) admissible extension such that $E \subset E'$. So $\exists a \in E' \setminus E$. We have previously established that $\exists \kappa \subseteq E \subset E'$ is such that $\kappa \triangleright a$, so E' is not weakly conflict-free, and therefore not strongly conflict-free. This is a contradiction with the fact that E' is a strong (respectively weak) admissible extension, and thus E is a \subseteq -maximal strong (respectively weak) admissible extension, *i.e.* a strong (respectively weak) preferred extension. ◀

► **Example 19.** Now we give the strong stable extensions of $\text{StrAF}_{\text{john}}$ and $\text{StrAF}_{\text{john}}^2$, with coval = \sum . It is easy to see that $\text{st}_S^\sum(\text{StrAF}_{\text{john}}) = \{\{B, D, F\}\}$ since B , D and F respectively defeat A , C and E ; this is the same result given by the "classical" defeat-based stable semantics (see Example 7), and it is the optimal solution for John. In the case of the second StrAF, we obtain $\text{st}_S^\sum(\text{StrAF}_{\text{john}}^2) = \{\{A, C, D, F\}, \{B, D, F\}\}$. We observe

that the optimal solution $\{A, C, D, F\}$ is now an extension, contrary to the outcome of the defeat-based stable semantics.

Since both StrAFs are symmetric, the weak stable extensions are equal to the strong stable extensions (see Proposition 29 for the proof), but it is not the case in general (see Example 25 for an illustration).

Now we prove that Dung's argumentation theory is an instance of our StrAF, as it is shown in the following result, that proves a one to one correspondence between the semantics of a Dung's AFs and the strong semantics of StrAFs.

► **Definition 20.** *Given an argumentation framework $AF = \langle \mathcal{A}, \mathcal{R} \rangle$, the StrAF associated with AF is $StrAF_{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ with $\mathcal{S}(a) = 1, \forall a \in \mathcal{A}$ and $\text{coval} = \sum$.*

► **Lemma 21.** *Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be a Dung's AF, and $StrAF_{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ its associated StrAF. The set $S \subseteq \mathcal{A}$ is conflict-free in AF if and only if it is strongly conflict-free in $StrAF_{AF}$.*

Lemma 21 is obvious from the definition of strong conflict-freeness.

► **Lemma 22.** *Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be a Dung's AF, and $StrAF_{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ its associated StrAF. The set $S \subseteq \mathcal{A}$ defends the argument $a \in \mathcal{A}$ in AF if and only if $S \subseteq \mathcal{A}$ defends the argument $a \in \mathcal{A}$ in $StrAF_{AF}$.*

Proof. Let us suppose that $S \subseteq \mathcal{A}$ defends the argument $a \in \mathcal{A}$ in AF . This means that for each $b \in \mathcal{A}$ such that $(b, a) \in \mathcal{R}$, $\exists c \in S$ such that $(c, b) \in \mathcal{R}$.

Now, let us consider an accrual $\kappa_1 \subseteq \mathcal{A}$ such that $\kappa_1 \triangleright a$. As established previously, $\forall b \in \kappa_1$, $\exists c \in S$ such that $(c, b) \in \mathcal{R}$, i.e. $\exists \kappa_2 = \{c\} \subseteq S$ such that κ_2 attacks b . Since $\text{coval}(\kappa_2) = \mathcal{S}(c) = 1 = \mathcal{S}(b)$, $\kappa_2 \triangleright b$ and thus $\kappa_2 \triangleright \kappa_1$. So S defends the argument a in $StrAF_{AF}$.

Now we suppose that $S \subseteq \mathcal{A}$ defends the argument $a \in \mathcal{A}$ in $StrAF_{AF}$, i.e. for all accruals κ_1 that defeat a , $\exists \kappa_2 \subseteq S$ such that $\kappa_2 \triangleright \kappa_1$. Since all arguments strengths are equal to 1, every argument $b \in \mathcal{A}$ attacking a corresponds to an accrual $\kappa_1 = \{b\}$ defeating a . So $\exists \kappa_2 \subseteq S$ such that $\kappa_2 \triangleright \kappa_1 = \{b\}$, and thus $\exists c \in \kappa_2 \subseteq S$ such that $(c, b) \in \mathcal{R}$. So we conclude that S defends the argument a in AF . ◀

► **Proposition 23.** (*Dung Compatibility*) *Let $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ be a Dung's AF, and $StrAF_{AF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$, with $\text{coval} = \sum$, its associated StrAF. For $\sigma \in \{\text{cf}, \text{ad}, \text{pr}, \text{st}\}$, $\sigma(AF) = \sigma_S(StrAF_{AF})$.*

Proof. The proof follows from Definition 17, Lemma 21 and Lemma 22. ◀

This result holds for any semantics based on conflict-freeness and defence, including those which are out of the scope of this paper. From the above, and the observation that for a StrAF associated with a Dung's theory AF , the notions of strong and weak conflict-freeness coincide, we easily obtain the following corollary.

► **Corollary 24.** *Let AF be a Dung's AF, and $StrAF_{AF}$ its associated StrAF. For $\sigma \in \{\text{cf}, \text{ad}, \text{pr}, \text{st}\}$, $\sigma(AF) = \sigma_W(StrAF_{AF})$.*

Proposition 23 and Corollary 24 are major tools for providing hardness results in the complexity study of StrAFs, in the next Section.

4 Complexity and Algorithms

This section discusses the complexity of various reasoning problems for StrAFs as well as algorithms for solving them. In the rest of the paper, we assume that coval is tractable, *i.e.* it can be computed in polynomial time. Usual aggregation operators (like \sum , Π or \max) are tractable. We suppose that the reader is familiar with the basic notions of complexity theory; for more details about this topic, we refer the reader to *e.g.* [6].

4.1 Complexity of Acyclic Frameworks

A $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ is *acyclic* if the relation \mathcal{R} is acyclic. Recall that Dung's acyclic argumentation frameworks have exactly one stable extension. The following example shows that this is not the case for StrAFs under the strong stable semantics.

► **Example 25.** Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a StrAF with $\mathcal{A} = \{a, b\}$, $\mathcal{R} = \{(a, b)\}$, $\mathcal{S}(a) = 1$, $\mathcal{S}(b) = 2$. For $\text{coval} = \sum$, we have $\text{st}_S^\sum(StrAF) = \emptyset$, whereas $\text{st}_W^\sum(StrAF) = \{\{a, b\}\}$.

In fact, we prove the existence of a weak stable extension for every acyclic StrAF, by providing Algorithm 1. Given $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ and an argument $a \in \mathcal{A}$, the set of attackers of a , denoted by $\Gamma_{StrAF}^-(a)$, is defined as $\Gamma_{StrAF}^-(a) = \{b \mid b \in \mathcal{A} \text{ and } (b, a) \in \mathcal{R}\}$. When $StrAF$ is clear from the context we write simply $\Gamma^-(a)$.

■ **Algorithm 1** Algorithm compute-acyclic-extension($\langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$, coval)

```

1  $E = \emptyset$ ;
2 while  $\mathcal{A} \neq \emptyset$  do
3    $E' = \{a \mid a \in \mathcal{A} \text{ such that } \Gamma^-(a) = \emptyset\}$ ;
4    $E = E \cup E'$ ;
5    $A' = \{a \mid a \in \mathcal{A}, \text{ and } \exists \kappa \subseteq E \text{ such that } \kappa \triangleright_{\text{coval}} a\}$ ;
6    $\mathcal{A} = \mathcal{A} \setminus (E \cup A')$ ;
7 end
8 return  $E$ ;

```

► **Proposition 26.** (*Termination and Correctness for Acyclic StrAFs*) Algorithm 1 compute-acyclic-extension always terminates. The set E returned by the algorithm is the unique weak stable extension of the input acyclic $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ with respect to coval.

Proof. Let us first prove that Algorithm 1 always terminates. The only modifications of the argumentation graph are removals of arguments (step 6). One can remark that removing nodes from an acyclic graph leads to another acyclic graph. Thus, at each iteration of the loop, there is at least one non-attacked argument a that is added to E' (step 3). This implies that each iteration of the loop strictly decreases the size of \mathcal{A} . After at most $|\mathcal{A}|$ iterations, \mathcal{A} is guaranteed to be empty, so the algorithm terminates.

Now let us prove that the algorithm is correct. Let E be the set returned by Algorithm 1 compute-acyclic-extension. Let us now suppose that E is not a weak stable extension of the input acyclic $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ with respect to coval. According to Definition 17, we can have two cases.

Case 1 E is not weakly conflict free. That means $\exists \kappa \subseteq E$ an accrual and $\exists b \in E$ such that $\kappa \triangleright b$. This is impossible: since b is defeated, it implies that $\Gamma^-(b) \neq \emptyset$, thus b is not

added to the sets E' and E (steps 3 and 4). Then, steps 5 and 6 exclude from \mathcal{A} the argument b , so it will never be added to E' at the following iterations of the `while` loop.

Case 2 There exists an argument $a \in \mathcal{A} \setminus E$ such that $\forall \kappa \subseteq E, \kappa \not\vdash a$. However, by construction, such an argument a is never added to A' (step 5), so it is not removed from \mathcal{A} at step 6. At some point, all its attackers must have been removed, so $\Gamma^-(a) = \emptyset$; this implies that a is added to E at some iteration of the loop (steps 3 and 4), which is a contradiction since by assumption $a \in \mathcal{A} \setminus E$.

Thus E is a weak stable extension.

Now let us prove that E is the unique weak stable extension. Let us suppose that $\exists E' \in \text{st}_W^{\text{coval}}(\text{StrAF})$ such that $E \neq E'$. We show iteratively that such an extension E' contains exactly the same arguments than E , which is contradictory.

First, we observe that the non-empty set U of arguments that are unattacked in StrAF is such that $U \subseteq E$ and $U \subseteq E'$. Indeed, these are the arguments that are added to E at the first iteration of the loop, step 3 and 4 of Algorithm 1. Since every unattacked argument a cannot be defeated by an accrual, it has to belong to every weak stable extension and in particular to E' . At this step, we observe that the set of arguments D that are defeated by any accrual $\kappa \subseteq U$ cannot belong to E' , since these arguments are defeated by some accrual in E' . These arguments are exactly the ones that are put aside from E by step 5 and 6 of Algorithm 1.

Since StrAF is assumed to be acyclic, arguments that are defeated by some accrual in D cannot be defeated by some accruals that are not in D , and are then straightforwardly defended by some accrual in U . Thereby these arguments must belong to E' since they are defended by E' and moreover cannot be defeated by any defended accrual. We also observe that this arguments are the one added to E at steps 3 and 4 of the next iteration of Algorithm 1 following steps 5 and 6 which has put aside arguments of D .

This process then repeats iteratively until it terminates since \mathcal{A} is finite and StrAF is acyclic. One can observe that, at each step of this process, arguments added to E by Algorithm 1 must be in E' , and conversely arguments set aside by Algorithm 1 cannot belong to E' . We thus finally obtain that $E = E'$ which contradicts our assumption, and so E is the unique weak stable extension. \blacktriangleleft

Algorithm 1 is a polynomial time procedure when the set A' can be computed in polynomial time.

► Proposition 27. (*Extension Computation for Acyclic StrAFs*) Let $\text{StrAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be an acyclic StrAF and coval an aggregation operator. Computing a weak stable extension of StrAF is polynomial.

Proof. As it is explained in the proof of Proposition 26, we know that the number of iterations of the `while` loop is at most $|\mathcal{A}|$. Determining the set E' of non-attacked arguments is doable in polynomial time, as well as the standard set operations at steps 4 and 6. Since coval is tractable, the computation of the set A' is also polynomial. This concludes the proof. \blacktriangleleft

We notice that Algorithm 1 corresponds to the well-known algorithm for computing the grounded extension in Dung's theory. This is not surprising, since the stable semantics coincide with the single-status grounded semantics for acyclic graphs (as well as any reasonable semantics).

Although we know that strong stable extensions may not exist for acyclic graphs, we show that deciding whether they exist or not is tractable.

► **Proposition 28.** (*Extension Existence for Acyclic StrAFs*) Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be an acyclic StrAF and coval an aggregation operator. Deciding whether StrAF has a strong stable extension is polynomial.

Proof. We know that if $StrAF$ admits a strong stable extension, then it is also a weak stable extension. Since $StrAF$ is acyclic, it admits a single weak stable extension, which is the set polynomially computed by Algorithm 1. This set E (by definition of weak stable extensions) defeats every argument $a \in \mathcal{A} \setminus E$. For deciding whether E is also the strong stable extension, we just have to check if it is strongly conflict-free; that can be done in polynomial time. ◀

Remark: Since there is a single weak stable extension, and computing this extension is polynomial (Proposition 27), both skeptical and credulous acceptance coincide for weak stable semantics, and are polynomial reasoning tasks. Similarly, Proposition 28 shows that, if a strong stable extension exists for an acyclic StrAF then it can be polynomially computed by Algorithm 1, and thus credulous (and skeptical) acceptance are polynomial as well.

4.2 Complexity of Symmetric Frameworks

Now we consider the case of symmetric StrAFs, *i.e.* frameworks for which $(a, b) \in \mathcal{R}$ if and only if $(b, a) \in \mathcal{R}$. This corresponds to situations like the one depicted in Example 5 with John's gardens problem. We prove here that, in spite of the specific structure of this subclass of StrAFs, reasoning is at the first level of the polynomial hierarchy (*i.e.* the same complexity as general StrAFs, as shown in Section 4.3).

First, we prove that strong and weak semantics coincide for symmetric StrAFs.

► **Proposition 29.** (*Weak/Strong Coincidence*) Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a symmetric StrAF and coval an aggregation operator. $\text{cf}_S^{\text{coval}}(StrAF) = \text{cf}_W^{\text{coval}}(StrAF)$.

Proof. We already know that $\text{cf}_S^{\text{coval}}(StrAF) \subseteq \text{cf}_W^{\text{coval}}(StrAF)$ (straightforward from the definition of strong and weak conflict-freeness). Now, let us prove the opposite inclusion, using *reductio ad absurdum*. We suppose that $\exists E \in \text{cf}_W^{\text{coval}}(StrAF)$ such that $E \notin \text{cf}_S^{\text{coval}}(StrAF)$. This means that $\exists a, b \in E$ such that $(a, b) \in \mathcal{R}$. Since $StrAF$ is symmetric, we know that $(b, a) \in \mathcal{R}$ also holds. Now, if $\mathcal{S}(a) \geq \mathcal{S}(b)$ then $\{a\} \triangleright_{\text{coval}} b$, otherwise $\mathcal{S}(b) > \mathcal{S}(a)$ and then $\{b\} \triangleright_{\text{coval}} a$. Whatever the situation, E is not weakly conflict-free, which is a contradiction. So we deduce that $\text{cf}_W^{\text{coval}}(StrAF) \subseteq \text{cf}_S^{\text{coval}}(StrAF)$, and we conclude $\text{cf}_S^{\text{coval}}(StrAF) = \text{cf}_W^{\text{coval}}(StrAF)$. ◀

Proposition 29 implies that $\text{ad}_S^{\text{coval}}(StrAF) = \text{ad}_W^{\text{coval}}(StrAF)$, $\text{pr}_S^{\text{coval}}(StrAF) = \text{pr}_W^{\text{coval}}(StrAF)$ and $\text{st}_S^{\text{coval}}(StrAF) = \text{st}_W^{\text{coval}}(StrAF)$, for any symmetric $StrAF$ and any coval .

Now we study credulous and skeptical reasoning; we show that deciding whether an argument belongs to some (respectively each) strong (or weak) extension is NP-complete (respectively CONP-complete).

► **Proposition 30.** (*Argument Acceptance for Symmetric StrAFs*) Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a symmetric StrAF and coval an aggregation function. Given $a \in \mathcal{A}$ and $X \in \{S, W\}$,

- deciding whether $\exists E \in \text{st}_X(StrAF)$ such that $a \in E$ is NP-complete.
- deciding whether $\forall E \in \text{st}_X(StrAF)$, $a \in E$ is CONP-complete.

Proof. It has been proven [38, 39] that credulous (respectively skeptical) acceptance is NP-complete (respectively coNP-complete) for symmetric argumentation frameworks. From Proposition 23 and Corollary 24, we know that any argumentation framework can be associated with a StrAF that has exactly the same set of (weak or strong) extensions. So, for a symmetric $AF = \langle \mathcal{A}, \mathcal{R} \rangle$ and $a \in \mathcal{A}$, we can define a symmetric StrAF such that its (weak or strong) stable extensions correspond to the stable extensions of AF . So, a is credulously (respectively skeptically) accepted in AF with respect to the stable semantics if and only if a is credulously (respectively skeptically) accepted in $StrAF$ with respect to (weak or strong) stable semantics. We can conclude that credulous (respectively skeptical) acceptance for StrAFs is NP-hard (respectively coNP-hard).

For NP-membership and coNP-membership, the proof is the same as for Proposition 35 in the case of general StrAFs. \blacktriangleleft

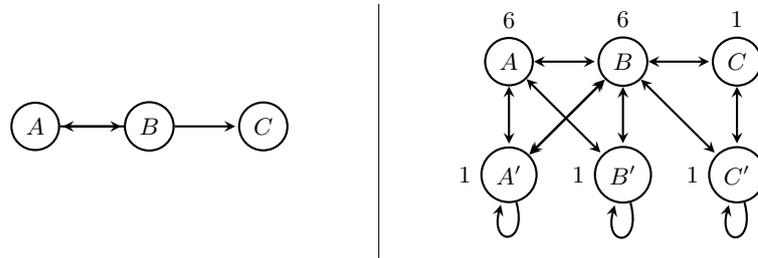
► **Proposition 31.** (*Extension Existence for Symmetric StrAFs*) Given $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ a symmetric StrAF and coval an aggregation function, checking whether $\text{st}_X^{\text{coval}}(StrAF) \neq \emptyset$ is NP-complete, with $X \in \{S, W\}$.

Proof. The NP-membership can be deduced from the fact the extension existence belongs to NP in the case of general StrAFs (see Proposition 34).

For proving NP-hardness, we use a reduction similar to the one proposed in [38]. Let us consider any $AF = \langle A, R \rangle$. We define a symmetric StrAF such that its (weak or strong) stable extensions exactly coincide with the stable extensions of AF . Formally, $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ with:

- $\mathcal{A} = A \cup A'$, where $A' = \{x' \mid x \in A\}$, i.e. for any argument $x \in A$, we create x' a copy of x ;
- $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3 \cup \mathcal{R}_4$, where
 - $\mathcal{R}_1 = \{(a, b), (b, a) \mid (a, b) \in R\}$, i.e. all the attacks from AF are made symmetric;
 - $\mathcal{R}_2 = \{(a, a'), (a', a) \mid a \in A\}$, i.e. each argument from AF attacks (and is attacked by) its copy;
 - $\mathcal{R}_3 = \{(a, b'), (b', a) \mid (a, b) \in R\}$, i.e. if a attacks b in AF , then a attacks (and is attacked by) the copy of b ;
 - $\mathcal{R}_4 = \{(a', a') \mid a' \in A'\}$, i.e. each copy is a self-attacking argument.
- if $\exists b \in A$ such that $(a, b) \in R$, then $\mathcal{S}(a) = M$, with M an arbitrary large number ($M \geq |A|$), otherwise $\mathcal{S}(a) = 1$.

Figure 5 illustrates the reduction from an AF to its corresponding symmetric StrAF.



■ **Figure 5** An example of our reduction from AF to $StrAF$

Let us prove that $\text{st}(AF) = \text{st}_S(StrAF)$. We start by the proof that $\text{st}(AF) \subseteq \text{st}_S(StrAF)$. Let $E \in \text{st}(AF)$ be a stable extension of AF . By definition, E is conflict-free in AF . Moreover, by the definition of $StrAF$, $\forall a, b \in A$, if $(a, b) \notin R$ and $(b, a) \notin R$, then $(a, b) \notin \mathcal{R}$

and $(b, a) \notin \mathcal{R}$: the construction of the *StrAF* makes the attacks in *AF* symmetrical, but no new conflicts are created. So E is strongly conflict-free in *StrAF*.

Now, we need to prove that each argument $a \in \mathcal{A} \setminus E$ is defeated by an accrual $\kappa \subseteq E$. Let us consider such an argument $a \in \mathcal{A} \setminus E$.

- If $a \in A \setminus E$ (i.e. a is an argument that belongs to *AF*, not a copy), then (since E is a stable extension of *AF*) $\exists b \in E$ such that $(b, a) \in \mathcal{R}$. This implies that $\mathcal{S}(b) = M$, so whatever the strength of a , $\mathcal{S}(b) \geq \mathcal{S}(a)$. This means that $\{b\} \triangleright a$, where $\{b\}$ is an accrual included in E .
- Otherwise, $a \in A'$. This means that $a = x'$, for some $x \in A$. Moreover, by definition of *StrAF*, $\mathcal{S}(a) = 1$.
 - If $x \in E$, since $(x, x') \in \mathcal{R}$, there is an accrual $\{x\} \subseteq E$ such that $\{x\} \triangleright x'$, i.e. $\{x\} \triangleright a$.
 - If $x \notin E$, $\exists y \in E$ such that $(y, x) \in \mathcal{R}$. This implies that $(y, x') \in \mathcal{R}$, thus $\{y\} \triangleright x'$, i.e. $\{y\} \triangleright a$.

We can conclude that $E \in \text{st}_S(\text{StrAF})$, so $\text{st}(\text{AF}) \subseteq \text{st}_S(\text{StrAF})$.

Now, we prove that $\text{st}_S(\text{StrAF}) \subseteq \text{st}(\text{AF})$. Let E be a strong stable extension of *StrAF*. First, we notice that there can be no argument $a' \in A'$ in the extension: each $a' \in A'$ attacks itself, so if $a' \in E$, then E is not strongly conflict-free, and thus E cannot be a strong stable extension. So $E \subseteq A$.

Since E is strongly conflict-free in *StrAF*, we can prove that E is conflict-free in *AF*. Indeed, the construction of *StrAF* from *AF* only adds attacks (by making the conflicts symmetrical). So, if there is no attack between two arguments $a, b \in A$ in *StrAF*, it means that there is no attack between them in *AF* either.

We still have to prove that E attacks each argument $b \in A \setminus E$ in *AF*. Using *reductio ad absurdum*, we suppose that $\exists b \in A \setminus E$ such that b is not attacked by E in *AF*, i.e. $\forall a \in E, (a, b) \notin \mathcal{R}$. This implies (by definition of *StrAF*) that, $\forall a \in E, (a, b') \notin \mathcal{R}$. So $\nexists \kappa \subseteq E$ such that $\kappa \triangleright b'$: the copy b' of b is not attacked, and thus not defeated by E in *StrAF*. So E is not a strong stable extension of *StrAF*: this is a contradiction.

We can now conclude that $E \in \text{st}(\text{AF})$, so $\text{st}_S(\text{StrAF}) \subseteq \text{st}(\text{AF})$.

We have proven that $\text{st}(\text{AF}) = \text{st}_S(\text{StrAF})$, which means that $\text{st}(\text{AF}) \neq \emptyset$ if and only if $\text{st}_S(\text{StrAF}) \neq \emptyset$. And it is known that verifying whether an argumentation framework has at least one stable extension is a NP-hard problem [39]. So we can conclude that verifying whether a symmetric StrAF has at least one strong stable extension is a NP-hard problem. The result also holds for weak stable semantics, since weak and strong semantics coincide for symmetric StrAFs. ◀

While verifying the existence of at least one (strong or weak) stable extension is generally hard, even for symmetric StrAFs, we show that this problem is trivial in the case of irreflexive symmetric StrAFs, i.e. symmetric StrAFs where no self-attack appear. Indeed, Algorithm 2 is a simple (non-deterministic) polynomial time procedure that returns a strong stable extension of the irreflexive symmetric StrAF given as input. This algorithm proves the existence of strong, and therefore weak as well, stable extensions for every irreflexive symmetric StrAF.

► **Proposition 32.** (*Termination and Correctness for Symmetric StrAFs*) Algorithm 2 compute-symmetric-extension always terminates. Any set E returned by the algorithm is a strong (and weak) stable extension of the irreflexive symmetric StrAF $= \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ with respect to coval.

Proof. Let us first prove that the algorithm always terminates. For any $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$, since $\forall a \in \mathcal{A}, \mathcal{S}(a) \in \mathbb{N}$, there is at least one argument such that $\mathcal{S}(a)$ is maximal. Exactly one of these arguments is selected at step 3, added to E , and removed from \mathcal{A} at step 5. This means that every iteration of the `while` loop strictly decreases the size of \mathcal{A} , thus the algorithm terminates in at most $|\mathcal{A}|$ iterations.

Now let us prove that the algorithm is correct. Let E be the set returned by some execution of Algorithm 2 *compute-symmetric-extension*. Let us now suppose that E is not a *strong stable extension* of the input symmetric $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$. According to Definition 17, we can have two cases.

Case 1 E is not *strongly conflict free*. That means $\exists a, b \in E$ such that $(a, b) \in \mathcal{R}$. This is impossible: without loss of generality, let us suppose that a has been added to E before b . Then, b is removed from \mathcal{A} at step 5, and it will never be added to E at the following iterations of the loop. Since the graph is symmetric, $(a, b) \in \mathcal{R}$ implies $(b, a) \in \mathcal{R}$, so we observe the same situation if b is added in E before a .

Case 2 There exists an argument $b \in \mathcal{A} \setminus E$ that is not defeated by an accrual in E , *i.e.* $\forall \kappa \subseteq E, \kappa \not\triangleright b$. There are again several possible situations.

1. b is not attacked by E , *i.e.* $\forall a \in E, (a, b) \notin \mathcal{R}$. Then b is never removed from \mathcal{A} at step 5, this implies that b will be added to E . This is a contradiction with the assumption that $b \in \mathcal{A} \setminus E$.
2. There are attacks from some arguments $a_1, \dots, a_k \in E$ towards b , such that $\text{coval}(\kappa) \not\geq \mathcal{S}(b)$, with $\kappa = \{a_1, \dots, a_k\}$. If $\exists a_i \in \kappa$ such that $\mathcal{S}(a_i) \geq \mathcal{S}(b)$, then $\kappa' = \{a_i\}$ is an accrual that defeats b , that is a contradiction with the assumption. So we can ensure that $\forall a_i \in \kappa, \mathcal{S}(a_i) < \mathcal{S}(b)$. In such a situation, b should have been selected and added to E before any argument a_i . This is again a contradiction with the assumption that $b \in \mathcal{A} \setminus E$.

Thus E is a *strong stable extension*, and therefore a *weak stable extension*. ◀

■ **Algorithm 2** Algorithm *compute-symmetric-extension*($\langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle, \text{coval}$)

```

1  $E = \emptyset$ ;
2 while  $\mathcal{A} \neq \emptyset$  do
3   | Select  $a \in \mathcal{A}$  s.t.  $\nexists a' \in \mathcal{A}$  with  $\mathcal{S}(a') > \mathcal{S}(a)$ ;
4   |  $E = E \cup \{a\}$ ;
5   |  $\mathcal{A} = \mathcal{A} \setminus (\{a\} \cup \{b \mid b \in \mathcal{A}, (a, b) \in \mathcal{R}\})$ ;
6 end
7 return  $E$ ;
```

4.3 General Complexity

Now we investigate the complexity of reasoning with general StrAFs. We first prove that verifying whether a set of arguments is a (strong or weak) stable extension is polynomial.

► **Proposition 33.** (*Extension Verification*) Let $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a StrAF and *coval* an aggregation function. For $E \subseteq \mathcal{A}$, checking whether $E \in \text{st}_X^{\text{coval}}(StrAF)$, with $X \in \{S, W\}$, is polynomial.

Proof. For $E \subseteq \mathcal{A}$, we have to check whether E is strongly (or weakly) conflict-free.

For the case of strong conflict-freeness, we have to check that for any two $a, b \in E$, it holds that $(a, b) \notin \mathcal{R}$. This is polynomial and corresponds to checking conflict-freeness in Dung's framework.

Weak conflict-freeness is verified by first computing the set $\Gamma^-(a) = \{b \mid b \in E \text{ such that } (b, a) \in \mathcal{R}\}$ for each $a \in E$, and then checking that it is not the case that $\text{coval}(\Gamma^-(a)) \geq \mathcal{S}(a)$. Since coval is assumed to satisfy the Accrual property (see Definition 11), $\text{coval}(\Gamma^-(a)) < \mathcal{S}(a)$ implies that for any accrual $\kappa \subseteq E$ that attacks a , $\text{coval}(\kappa) \leq \text{coval}(\Gamma^-(a)) < \mathcal{S}(a)$, so $\kappa \not\vdash a$. All the above tests can be carried out in polynomial time.

Finally, for both weak and strong semantics the following check is needed. For every $a \notin E$, the set $\Gamma^-(a) = \{b \mid b \in E \text{ such that } (b, a) \in \mathcal{R}\}$ is computed. Then for each such a it is verified that $\text{coval}(\Gamma^-(a)) \geq \mathcal{S}(a)$. Again, these tests are computable in polynomial time. ◀

Computing stable extensions is a central problem in argumentation. Similar to the case of Dung's theories, this problem is intractable for StrAFs. However, its complexity does not increase for StrAFs.

► **Proposition 34.** (*Extension Existence*) *Given StrAF = $\langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ a StrAF and coval an aggregation function, checking whether $\text{st}_X^{\text{coval}}(\text{StrAF}) \neq \emptyset$ is NP-complete, with $X \in \{S, W\}$.*

Proof. It is known that verifying whether a Dung's AF has at least one stable extension is NP-complete [39]. From Proposition 23 and Corollary 24, we know that any AF can be associated with a StrAF that has exactly the same set of (strong or weak) extensions. We can deduce that checking the existence of (strong or weak) stable extensions of a StrAF is at least as hard as checking the existence of stable extensions of an AF. So we conclude that it is NP-hard.

NP-membership follows from the polynomial complexity of Verification (see Proposition 33). ◀

Finally, we show that, likewise Dung's AFs, credulous and skeptical acceptance are at the first level of the polynomial hierarchy.

► **Proposition 35.** (*Argument Acceptance*) *Let StrAF = $\langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ be a StrAF and coval an aggregation function. Given $a \in \mathcal{A}$,*

- *deciding whether $\exists E \in \text{st}_X(\text{StrAF})$ such that $a \in E$ is NP-complete.*
- *deciding whether $\forall E \in \text{st}_X(\text{StrAF}), a \in E$ is coNP-complete.*

Proof. Since credulous (respectively skeptical) argument acceptance is NP-hard (respectively coNP-hard) for the specific case of symmetric StrAFs (see Proposition 30), it is also NP-hard (respectively coNP-hard) in the general case.

For NP-membership of credulous acceptance, let $E \subseteq \mathcal{A}$ be a set of arguments such that $a \in E$. Verifying whether E is a strong (or weak) stable extension is polynomial (Proposition 33); the conclusion follows.

For coNP-membership of skeptical acceptance, let $E \subseteq \mathcal{A}$ be a set of arguments such that $a \notin E$. Again, verifying whether E is a strong (or weak) stable extension is polynomial; this concludes the proof. ◀

Computing the stable extensions of a StrAF is a challenging problem even when the aggregation function is $\text{coval} = \sum$. The rest of this section presents an algorithm that finds the stable extensions of $\text{StrAF} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$, with \sum as the aggregation function, by translating their computation into a *pseudo-Boolean constraint satisfaction* problem.

A pseudo-Boolean constraint is an inequality of the form $\sum_i w_i v_i \geq k$, where w_i and k are positive integers, and v_i is a variable that may assume the values 0 and 1. Nowadays, there are several systems that can solve pseudo-Boolean constraint satisfaction problems (e.g. SAT4J [47], OpenWBO [52], RoundingSat [41]), many of which draw on the power of efficient SAT solving techniques.

4.3.0.1 Weak Semantics

We start with weak stable semantics. Given $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ and $\text{coval} = \sum$, we associate a pseudo-Boolean constraint satisfaction problem $CS_W(StrAF) = (X, C)$, defined as follows:

- with each argument $a_i \in \mathcal{A}$ we associate a Boolean variable $x_i \in X$;
- the set of constraints C is as follows:
 - (1) For each $a \in \mathcal{A}$ with $\Gamma^-(a) = \{a_1, a_2, \dots, a_n\}$, add the constraint $\mathcal{S}(a_1) \times x_1 + \mathcal{S}(a_2) \times x_2 + \dots + \mathcal{S}(a_n) \times x_n < \mathcal{S}(a) \times x + (1-x) \times M^a$, where x is the variable that corresponds to a , and M^a an integer that is explained below.
 - (2) For each $a \in \mathcal{A}$ with $\Gamma^-(a) = \{a_1, a_2, \dots, a_n\}$, add the constraint $\mathcal{S}(a_1) \times x_1 + \mathcal{S}(a_2) \times x_2 + \dots + \mathcal{S}(a_n) \times x_n \geq (1-x) \times \mathcal{S}(a)$, where x is the variable that corresponds to a .

Intuitively, x_i is true means that the corresponding argument a_i belongs to an extension. Each solution of the problem corresponds then to one extension of the StrAF.

The value M^a that is associated with an argument $a \in \mathcal{A}$ in constraint (1) is an integer such that $M^a > \sum_{a_i \in \Gamma^-(a)} \mathcal{S}(a_i)$. The purpose that M^a serves is to “neutralize” constraint (1) when the variable x receives the value 0. Indeed, depending on the value of x , constraint (1) can be understood as follows.

- (1.1) if $x = 0$, the constraint becomes $\sum_{a_i \in \Gamma^-(a)} \mathcal{S}(a_i) \times x_i < M^a$, which is satisfied regardless of the values assigned to the variables x_i , since $\sum_{a_i \in \Gamma^-(a)} \mathcal{S}(a_i) \times x_i \leq \sum_{a_i \in \Gamma^-(a)} \mathcal{S}(a_i)$ and by construction $\sum_{a_i \in \Gamma^-(a)} \mathcal{S}(a_i) < M^a$. Therefore, in this case constraint (1) is inactive.
- (1.2) if $x = 1$, the constraint becomes $\sum_{a_i \in \Gamma^-(a)} \mathcal{S}(a_i) \times x_i < \mathcal{S}(a)$, requiring that $\text{coval}(\Gamma^-(a)) < \mathcal{S}(a)$.

In other words, constraint (1) means that we can accept a only if it is stronger than its accepted attackers; roughly speaking, these attackers form an accrual that attacks a but cannot defeat it even collectively.

Constraint (2) can be understood as follows:

- (2.1) if $x = 0$, the constraint becomes $\mathcal{S}(a_1) \times x_1 + \mathcal{S}(a_2) \times x_2 + \dots + \mathcal{S}(a_n) \times x_n \geq \mathcal{S}(a)$, i.e. $\text{coval}(\Gamma^-(a)) \geq \mathcal{S}(a)$.
- (2.2) if $x = 1$, the constraint becomes $\mathcal{S}(a_1) \times x_1 + \mathcal{S}(a_2) \times x_2 + \dots + \mathcal{S}(a_n) \times x_n \geq 0$, that is trivially satisfied; the constraint is inactive in this case.

Intuitively, this constraint can be interpreted as the fact that we can reject a only if its accepted attackers can join their respective strength to collectively overtake the strength of a . Roughly speaking, this means that there exists an accrual that defeats a .

Note that, when a has no attacker, constraint (2) becomes $0 \geq (1-x) \times \mathcal{S}(a)$, that can be rewritten into $x \geq 1$. This means that non-attacked arguments must be accepted.

The next proposition shows a direct correspondence between weak stable extensions of $StrAF$ and the solutions of the problem $CS_W(StrAF)$.

► **Definition 36.** Given a set of arguments $E \subseteq \mathcal{A}$, the Boolean assignment ω_E is defined, $\forall x_i \in X$, by $\omega_E(x_i) = 1$ if and only if $a_i \in E$.

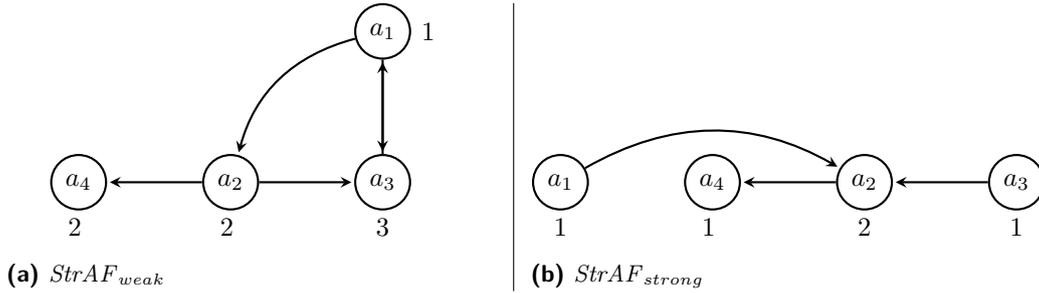
► **Proposition 37.** (Computation for Weak Stable Semantics) A set $E \subseteq \mathcal{A}$ is a weak stable extension of $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ with respect to \sum if and only if the assignment ω_E is a solution for $CS_W(StrAF)$.

► **Example 38.** Consider $StrAF_{weak} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ described at Figure 6a. The numerical values close to the nodes represent the strength of arguments *i.e.* the function $\mathcal{S}(\cdot)$. We observe that this StrAF does not admit a strong stable extension, but weak stable extensions can be computed as follows.

If we set M^a to 5 in all the constraints for simplicity, we obtain the following set of constraints:

$$\begin{aligned} 3 \times x_3 &< 1 \times x_1 + (1 - x_1) \times 5 \\ 1 \times x_1 &< 2 \times x_2 + (1 - x_2) \times 5 \\ 1 \times x_1 + 2 \times x_2 &< 3 \times x_3 + (1 - x_3) \times 5 \\ 2 \times x_2 &< 2 \times x_4 + (1 - x_4) \times 5 \\ 3 \times x_3 &\geq (1 - x_1) \times 1 \\ 1 \times x_1 &\geq (1 - x_2) \times 2 \\ 1 \times x_1 + 2 \times x_2 &\geq (1 - x_3) \times 3 \\ 2 \times x_2 + &\geq (1 - x_4) \times 2 \end{aligned}$$

The above set of constraints has the solutions $S_1 = \{x_1 = 1, x_2 = 1, x_3 = 0, x_4 = 0\}$, and $S_2 = \{x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0\}$ that correspond to $st_{\sum}^{WF}(StrAF_{weak}) = \{\{a_1, a_2\}, \{a_2, a_3\}\}$.



■ **Figure 6** Two Examples of StrAFs

4.3.0.2 Strong Semantics

For strong stable semantics, we use the encoding for weak stable semantics as a starting point. We only need to add a constraint for enforcing the strong conflict-freeness. Given $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ and $coval = \sum$, we associate a pseudo-boolean constraint satisfaction problem $CS_S(StrAF) = (X, C)$ as follows:

- with each argument $a_i \in \mathcal{A}$ we associate a Boolean variable $x_i \in X$;
- the set of constraints C is as follows:
 - (1) For each $a \in \mathcal{A}$ with $\Gamma^-(a) = \{a_1, a_2, \dots, a_n\}$, add the constraint $\mathcal{S}(a_1) \times x_1 + \mathcal{S}(a_2) \times x_2 + \dots + \mathcal{S}(a_n) \times x_n < \mathcal{S}(a) \times x + (1 - x) \times M^a$, where x is the variable that corresponds to a , and M^a an integer that is explained previously.
 - (2) For each $a \in \mathcal{A}$ with $\Gamma^-(a) = \{a_1, a_2, \dots, a_n\}$, add the constraint $\mathcal{S}(a_1) \times x_1 + \mathcal{S}(a_2) \times x_2 + \dots + \mathcal{S}(a_n) \times x_n \geq (1 - x) \times \mathcal{S}(a)$, where x is the variable that corresponds to a .
 - (3) For each pair of arguments $a_i, a_j \in \mathcal{A}$ such that $(a_i, a_j) \in \mathcal{R}$, add the constraint $x_i + x_j \leq 1$.

Constraints (1) and (2) have already been explained. Constraint (3) means that two conflicting arguments cannot be jointly accepted: this enforces (strong) conflict-freeness.

The next proposition shows that there is a direct correspondence between strong stable extensions of $StrAF$ and the solutions of the problem $CS_S(StrAF)$.

► **Proposition 39.** (*Computation for Strong Stable Semantics*) *A set $E \subseteq \mathcal{A}$ is strong stable extension of $StrAF = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ with respect to \sum if and only if the assignment ω_E is a solution of $CS_S(StrAF)$.*

► **Example 40.** Consider $StrAF_{strong} = \langle \mathcal{A}, \mathcal{R}, \mathcal{S} \rangle$ described at Figure 6b. We observe that if we use the classical defeat-based semantics, neither a_1 nor a_3 can defeat a_2 , while our semantics sensitive to accruals give a different result. The computation of strong stable extensions is made as follows. We translate the StrAF into a set of pseudo-boolean constraints.

$$\begin{array}{l} 0 < 1 \times x_1 + (1 - x_1) \times 5 \\ 1 \times x_1 + 1 \times x_3 < 2 \times x_2 + (1 - x_2) \times 5 \\ 0 < 1 \times x_3 + (1 - x_3) \times 5 \\ 2 \times x_2 < 1 \times x_4 + (1 - x_4) \times 5 \end{array} \quad \left| \begin{array}{l} 0 \geq (1 - x_1) \times 1 \\ 1 \times x_1 + 1 \times x_3 \geq (1 - x_2) \times 2 \\ 0 \geq (1 - x_3) \times 1 \\ 2 \times x_2 \geq (1 - x_4) \times 1 \end{array} \right. \quad \left| \begin{array}{l} x_1 + x_2 \leq 1 \\ x_3 + x_2 \leq 1 \\ x_2 + x_4 \leq 1 \end{array} \right.$$

The three columns correspond respectively to constraints (1), (2) and (3). The only solution of this set of constraints is $S = \{x_1 = x_3 = x_4 = 1, x_2 = 0\}$, that corresponds to the unique strong stable extension $E = \{a_1, a_3, a_4\}$.

We notice that $StrAF_{weak}$ from Example 38 does not have any strong stable extension. Indeed, none of the solutions S_1 and S_2 that we exhibited previously satisfies the strong conflict-freeness constraint.

5 Related Work

Several approaches have been proposed in the AI literature to model the accrual of arguments. Especially, in structured argumentation setting, [62, 49, 50, 44, 63] deal with accruals by adding a new argument that represents the accrual; these works consider that the arguments members of an accrual should not be taken into consideration on their own. On the other hand, [66] defines a formalism in which accruals are explicitly given, since conflicts are defined in terms of sets of arguments attacking other (sets of) arguments. The notions introduced in these works strongly rely on the internal structure of the arguments, whereas in abstract argumentation, which is the subject matter of our study, this structure is unknown.

Collective attacks have also been studied in abstract argumentation. In [57, 42] the authors proposed a generalization of Dung's abstract framework by presenting an abstract attack relation between sets of arguments and by extending the associated semantics. Similarly, in [23], n -ary conflicts are defined between sets of arguments that cannot be jointly accepted, while no explicit conflicts exist between these arguments. Besides, [19] defines collective argumentation frameworks where sets of arguments attacks sets of arguments. In [43, 67], combined attacks (*i.e.* several arguments attacking a same target) are modelled through notions of joint attack or accrual patterns. In these works there is no notion of arguments strength (and thus, defeat relation) and then compensation as we do here.

In all the works cited previously (in both fields of structured and abstract argumentation) accruals of arguments need to be explicitly elicited as an input of the reasoning process. In other words, each existing accruals must be clearly identified before the reasoning step of the process, and all combinations of arguments must be taken into account when a new argument is provided to the system. On the contrary, in our work accruals appear as an output of

the system, and are computed only from the attack relation and the arguments strength. Moreover, since in our work the elicitation of all accruals and preferences between sets of arguments is not necessary, our framework is more modular. Indeed, when new arguments are added to the framework or agents preferences are updated, the generation of new accruals or the loss of old ones is automatically handled by the behaviour of our semantics.

Another framework representing synergies of arguments is proposed in [45]. This framework proposes an extension of the value-based argumentation framework [10] by defining a defeat relation with varied strength. The strength of defeat of a subset A of arguments over another subset B of arguments depends then on the values promoted by A and B . This framework is also different to ours as in our framework the strength is associated with arguments and the collective strength of an accrual is computed through an aggregation operator that aggregates according to different ways the strengths of the individual arguments forming the accrual. Then this collective strength is used in the definition of a collective defeat relation that is used between subsets of arguments.

Abstract Dialectical Frameworks (ADFs) [21] associate to each abstract node x (called statement) an acceptance formula ϕ_x built on variables that correspond to the parents of the node x in the graph. Such an acceptance formula can model collective attacks, *e.g.* a and b collectively attack c is modelled by $\phi_c = \neg a \vee \neg b$: if both a and b are accepted, then c is rejected (because ϕ_c is false), otherwise c is accepted. However, ADFs cannot represent the notion of compensation between weak and strong arguments as we do. Moreover, the problem of explicit elicitation of collective defeats is the same as previously described. Although [22] extends ADFs by adding a notion of weight, it does not at all correspond to the strength of the arguments, but rather to the computation of a degree of acceptance for statements (in the spirit of gradual semantics [24]).

The idea of “collective support” between arguments is also explored in [26], where arguments participating in a coalition support (or help) each other against attacks from other arguments. However, this idea of an accrual of arguments with other arguments, especially when related to the notion of support, is quite different from our idea of “accrual” and “collective attack”. In our work, we consider that an accrual of arguments defends a common position on a matter which is defended individually by each of the participating arguments (by expressing a different point of view). The motivation of an accrual is not to create a mutual support among the participating arguments. We consider that this is rather closer to the notion of “extension”. In our work, arguments form an accrual in order to defeat a common adversary that none of them can defeat alone. Thus, the accrual forms a compound unified attack incorporating different points of view (or dimensions) against a single argument or a set of arguments defending a conflicting position.

We have mentioned previously that several works attach a weight to arguments (see *e.g.* [48, 9, 2, 8, 55, 61]). The meaning of these weights can be the votes of users, some notion of trust, or the certainty about the arguments premises. However, all these works strongly differ from ours since they intend to define a ranking or an acceptance degree of arguments. They focus on the individual acceptance of arguments, while the extension-based semantics that we define in this paper correspond to a notion of joint acceptance. Moreover, these works do not study collective defeat and accrual of arguments.

Recently, in [37, 31, 29] the authors investigated aggregation of weights in weighted argumentation systems. However, in this framework weights are associated with attack relations and are not involved in the definition of a defeat relation applied for accruals like in our framework. Moreover, none of the above works has investigated in depth the theoretical and computational properties of the proposed approaches for compound attacks. Similarly to

both papers cited above, [51] defines attacks of different strength. The paper generalizes the notion of defense by setting conditions about the relative strength of attackers and defenders in order to generalize Dung’s semantics. The concepts of collective attacks or accrual of arguments are not studied there. Finally, [16] defines a notion of collective attack by the aggregation of weights associated with these attacks. In this framework, if both a and b attack an argument c , respectively with strength m and n , then the strength of the collective attack from $\{a, b\}$ to c is the result of aggregating m and n (*e.g.* $m + n$). Nevertheless, this notion of collective attack is there used to relax the notion of defence, and this work does not address the notion of individual/collective attack/defeat like our approach.

6 Conclusion and Future Work

In this paper, we have proposed an intuitive framework for representing strength of arguments called Strength-based Argumentation Framework. We use the numerical strength of arguments to define extension-based semantics based on a defeat relation, that is the combination of the initial attacks between arguments and the comparison of their respective strengths. We have shown how quantitative strengths can be aggregated when several arguments attack a same target, making an *accrual of arguments*. The definition of collective defeat has allowed us to define new semantics that lead to interesting extensions that cannot be obtained when the usual Dung’s AFs or individual defeat are used.

We have established the complexity of several reasoning problems for special cases (acyclic and symmetric graphs) as well as general StrAFs. Surprisingly, we notice that, while our framework can model situations that are not captured by Dung’s AFs without a blow up of the framework size, the complexity of reasoning does not increase. Table 2 summarizes our results for (strong and weak) stable semantics. P stands for polynomial, while X-c means X-complete where X is either NP or coNP. We say that a problem is trivial when the answer is straightforwardly "yes". We recall that all these results hold under the assumption that coval can be computed in polynomial time.

	Acyclic		Symmetric		General	
	Strong	Weak	Strong	Weak	Strong	Weak
Verification	P	P	P	P	P	P
Existence	P	Trivial	NP-c	NP-c	NP-c	NP-c
Credulous	P	P	NP-c	NP-c	NP-c	NP-c
Skeptical	P	P	coNP-c	coNP-c	coNP-c	coNP-c

■ **Table 2** Summary of complexity results for (strong and weak) stable semantics

Our complexity results only concern the (strong and weak) stable semantics. For the case of weak semantics, they remind the complexity results for AFs with collective attacks [40]. We plan to deepen this investigation and consider other semantics as well, especially for the strong versions since the weak versions can be deduced from [40]. Moreover, regarding symmetric StrAFs, we have proven that extension existence is a trivial problem if we only consider irreflexive attack relations, *i.e.* no self-attacking arguments belong to the StrAF. It is known that reasoning with irreflexive symmetric AFs is polynomial [28]. We will investigate this question for irreflexive StrAFs, and determine whether self-attacks are the source of NP (or coNP) hardness for symmetric StrAFs, as they are in the case of symmetric AFs.

Weak extensions make sense for applications where conflicts are expected between pieces of information, but these pieces of information appear to be not strong enough to actually

exclude each other. In other words, evaluation of these conflicts does not lead to reject some of the involved pieces of information which thus can be accepted together. However, in the case of logic-based argumentation, the loss of (strong) conflict-freeness may lead to problems with the inference defined from the argumentation framework if two arguments belong to the same extension while one of them undercuts the other one. We plan to refine the notion of weak conflict-freeness according to this idea, similarly to the refinement of Preference-based AFs [5].

In Section 4.1, we have stated that acyclic StrAFs have exactly one (strong or weak) stable extension. This is reminiscent to known results for classical AFs [35], or argumentation frameworks with weighted attacks [17, 18]. These papers prove that (in these frameworks) the four classical semantics introduced by Dung (namely grounded, complete, preferred and stable) coincide for well-founded frameworks. The formal definition and study of the relations between these semantics in our framework is a promising track for future research.

We plan to study the logical properties of our accrual-based semantics with respect to the properties of coval operators. Different properties may be suited to different applications scenarios, thus indicating which coval operators to choose depending on the application domain or the nature of arguments.

As mentioned in Section 3.1, we have focused in this work on situations where arguments strengths are commensurable from an argument to another. Again, this assumption allows to directly compare strengths associated with different arguments, and it thus makes sense to then aggregate them to compute the strength associated with an accrual. However, this assumption may appear to be too strong for many interesting real-world applications, for example when weights represent subjective opinions or judgments, or when they are provided by different sources which do not share the meaning they associate with weights. We will investigate strategies adapted to applications where such commensurability assumption over strengths associated with arguments is dropped.

Lastly, frameworks where arguments are built from logical formulas or rules have received a lot of attention [46, 13, 36]. On the other hands, some logical frameworks allow representation of weighted pieces of information [58, 34]. A natural question is then the generation of structured arguments from such weighted logics, and then whether and how such structured arguments can be combined in accruals.

References

- 1 Leila Amgoud, Jonathan Ben-Naim, Dragan Doder, and Srdjan Vesic. Ranking arguments with compensation-based semantics. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning, KR'16*, pages 12–21, 2016.
- 2 Leila Amgoud, Jonathan Ben-Naim, Dragan Doder, and Srdjan Vesic. Acceptability semantics for weighted argumentation frameworks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI'17*, pages 56–62, 2017.
- 3 Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.
- 4 Leila Amgoud and Henri Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3-4):413–436, 2009.
- 5 Leila Amgoud and Srdjan Vesic. Rich preference-based argumentation frameworks. *International Journal of Approximate Reasoning*, 55(2):585–606, 2014.
- 6 Sanjeev Arora and Boaz Barak. *Computational Complexity - A Modern Approach*. Cambridge University Press, 2009.
- 7 Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. Abstract argumentation frameworks and their semantics. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and

- Leendert van der Torre, editors, *Handbook of Formal Argumentation*, pages 159–236. College Publications, 2018.
- 8 Pietro Baroni, Antonio Rago, and Francesca Toni. How many properties do we need for gradual argumentation? In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI'18*, pages 1736–1743, 2018.
 - 9 Pietro Baroni, Marco Romano, Francesca Toni, Marco Aurisicchio, and Giorgio Bertanza. Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation*, 6(1):24–49, 2015.
 - 10 Trevor Bench-Capon. Value-based argumentation frameworks. In *Proceedings of the Ninth International Workshop on Non-Monotonic Reasoning, NMR'02*, pages 443–454, 2002.
 - 11 Trevor Bench-Capon, Henry Prakken, and Giovanni Sartor. Argumentation in legal reasoning. In *Argumentation in Artificial Intelligence*, pages 363–382. 2009.
 - 12 Salem Benferhat, Didier Dubois, and Henri Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence, UAI'93*, pages 411–419, 1993.
 - 13 Philippe Besnard and Anthony Hunter. *Elements of Argumentation*. MIT Press, 2008.
 - 14 Stefano Bistarelli and Fabio Gadducci. Enhancing constraints manipulation in semiring-based formalisms. In *Proceedings of the Seventeenth European Conference on Artificial Intelligence, ECAI'06*, pages 63–67, 2006.
 - 15 Stefano Bistarelli, Daniele Pirolandi, and Francesco Santini. Solving weighted argumentation frameworks with soft constraints. In *Proceedings of the Twenty-Fifth Italian Conference on Computational Logic, CILC'10*, 2010.
 - 16 Stefano Bistarelli, Fabio Rossi, and Francesco Santini. A novel weighted defence and its relaxation in abstract argumentation. *International Journal of Approximate Reasoning*, 92:66–86, 2018.
 - 17 Stefano Bistarelli and Francesco Santini. Some thoughts on well-foundedness in weighted abstract argumentation. In *Proceedings of the Sixteenth International Conference on Principles of Knowledge Representation and Reasoning, KR'18*, pages 623–624, 2018.
 - 18 Stefano Bistarelli and Francesco Santini. Well-foundedness in weighted argumentation frameworks. In *Proceedings of the Sixteenth European Conference on Logics in Artificial Intelligence, JELIA'19*, pages 69–84, 2019.
 - 19 Alexander Bochman. Collective argumentation and disjunctive logic programming. *Journal of Logic and Computation*, 13(3):405–428, 2003.
 - 20 Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, and Nicolas Maudet. A comparative study of ranking-based semantics for abstract argumentation. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16*, pages 914–920, 2016.
 - 21 Gerhard Brewka, Hannes Strass, Stefan Ellmauthaler, Johannes Peter Wallner, and Stefan Woltran. Abstract dialectical frameworks revisited. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, IJCAI'13*, pages 803–809, 2013.
 - 22 Gerhard Brewka, Hannes Strass, Johannes Peter Wallner, and Stefan Woltran. Weighted abstract dialectical frameworks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI'18*, pages 1779–1786, 2018.
 - 23 Martin Caminada and Srdjan Vesic. On extended conflict-freeness in argumentation. In *Proceedings of the Twenty-Fourth Benelux Conference on Artificial Intelligence, BNAIC'12*, 2012.
 - 24 Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.
 - 25 Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Proceedings of the Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'05*, pages 378–389, 2005.

- 26 Claudette Cayrol and Marie-Christine Lagasque-Schiex. Coalitions of arguments: A tool for handling bipolar argumentation frameworks. *International Journal of Intelligent Systems*, 25(1):83–109, 2010.
- 27 Oana Cocarascu, Antonio Rago, and Francesca Toni. Extracting dialogical explanations for review aggregations with argumentative dialogical agents. In *Proceedings of the Eighteenth International Conference on Autonomous Agents and MultiAgent Systems, AAMAS'19*, pages 1261–1269, 2019.
- 28 Sylvie Coste-Marquis, Caroline Devred, and Pierre Marquis. Symmetric argumentation frameworks. In *Proceedings of the Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'05*, pages 317–328, 2005.
- 29 Sylvie Coste-Marquis, Sébastien Konieczny, Pierre Marquis, and Mohand Akli Ouali. Selecting extensions in weighted argumentation frameworks. In *Proceedings of the Fourth International Conference on Computational Models of Argument, COMMA'12*, pages 342–349, 2012.
- 30 Sylvie Coste-Marquis, Sébastien Konieczny, Pierre Marquis, and Mohand Akli Ouali. Weighted attacks in argumentation frameworks. In *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning, KR'12*, 2012.
- 31 Sylvie Coste-Marquis, Sébastien Konieczny, Pierre Marquis, and Mohand Akli Ouali. Weighted attacks in argumentation frameworks. In *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning, KR'12*, 2012.
- 32 Célia da Costa Pereira, Andrea Tettamanzi, and Serena Villata. Changing one's mind: Erase or rewind? In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI'11*, pages 164–171, 2011.
- 33 Yannis Dimopoulos, Jean-Guy Mailly, and Pavlos Moraitis. Argumentation-based negotiation with incomplete opponent profiles. In *Proceedings of the Eighteenth International Conference on Autonomous Agents and MultiAgent Systems, AAMAS'19*, pages 1252–1260, 2019.
- 34 Didier Dubois and Henri Prade. Possibilistic logic - an overview. In *Computational Logic*, pages 283–342. 2014.
- 35 Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
- 36 Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. Assumption-based argumentation. In *Argumentation in Artificial Intelligence*, pages 199–218. 2009.
- 37 Paul E. Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, 175(2):457 – 486, 2011.
- 38 Wolfgang Dvorák. *Computational Aspects of Abstract Argumentation*. PhD thesis, Technische Universität Wien, 2012. URL: <http://permalink.obvsg.at/AC07812708>.
- 39 Wolfgang Dvorák and Paul E. Dunne. Computational problems in formal argumentation and their complexity. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of Formal Argumentation*, pages 631–688. College Publications, 2018.
- 40 Wolfgang Dvorák, Alexander Greßler, and Stefan Woltran. Evaluating setafs via answer-set programming. In *Proceedings of the Second International Workshop on Systems and Algorithms for Formal Argumentation, SAFA'18*, pages 10–21, 2018.
- 41 Jan Elffers and Jakob Nordström. Divide and conquer: Towards faster pseudo-boolean solving. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI'18*, pages 1291–1299, 2018.
- 42 Giorgos Flouris and Antonis Bikakis. A comprehensive study of argumentation frameworks with sets of attacking arguments. *International Journal of Approximate Reasoning*, 109:55–86, 2019.
- 43 Dov M. Gabbay. Semantics for higher level attacks in extended argumentation frames part 1: Overview. *Studia Logica*, 93(2-3):357–381, 2009.
- 44 Thomas Gordon. Defining argument weighing functions. *Journal of Applied Logics - IfCoLog Journal of Logics and their Application*, 5:747–773, 2018.

- 45 Souhila Kaci and Christophe Labreuche. Representing synergy among arguments with choquet integral. In *Proceedings of the Twelfth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'13*, pages 302–314, 2013.
- 46 Antonis C. Kakas and Pavlos Moraitis. Argumentation based decision making for autonomous agents. In *Proceedings of the Second International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS'03*, pages 883–890, 2003.
- 47 Daniel Le Berre and Anne Parrain. The sat4j library, release 2.2. *Journal on Satisfiability, Boolean Modeling and Computation*, 7(2-3):59–6, 2010.
- 48 João Leite and João Martins. Social abstract argumentation. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI'11*, pages 2287–2292, 2011.
- 49 Mauro Javier Gómez Lucero, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. On the accrual of arguments in defeasible logic programming. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence, IJCAI'09*, pages 804–809, 2009.
- 50 Mauro Javier Gómez Lucero, Carlos Iván Chesñevar, and Guillermo Ricardo Simari. Modelling argument accrual with possibilistic uncertainty in a logic programming setting. *Information Sciences*, 228:1–25, 2013.
- 51 Diego C. Martínez, Alejandro Javier García, and Guillermo Ricardo Simari. An abstract argumentation framework with varied-strength attacks. In *Proceedings of the Eleventh International Conference on Principles of Knowledge Representation and Reasoning, KR'08*, pages 135–144, 2008.
- 52 Ruben Martins, Vasco Manquinho, and Inês Lynce. Open-wbo: A modular maxsat solver. In *Proceedings of the Seventeenth International Conference on Theory and Applications of Satisfiability Testing, SAT'14*, pages 438–445, 2014.
- 53 Paul-Amaury Matt and Francesca Toni. A game-theoretic measure of argument strength for abstract argumentation. In *Proceedings of the Eleventh European Conference on Logics in Artificial Intelligence, JELIA'08*, pages 285–297, 2008.
- 54 Peter McBurney, Simon Parsons, and Iyad Rahwan, editors. *Proceedings of the Eighth International Workshop on Argumentation in Multi-Agent Systems, ArgMAS'11*, volume 7543 of *Lecture Notes in Computer Science*. Springer, 2012.
- 55 Till Mossakowski and Fabian Neuhaus. Modular semantics and characteristics for bipolar weighted argumentation graphs. *CoRR*, abs/1807.06685, 2018.
- 56 Gopal Shyam Nair. How do reasons accrue? In E. Lord and B. Maguire, editors, *Weighing reasons*. Oxford University Press, 2016.
- 57 Søren Holbech Nielsen and Simon Parsons. A generalization of dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Proceedings of the Third International Workshop on Argumentation in Multi-Agent Systems, ArgMAS'06*, pages 54–73, 2006.
- 58 Gadi Pinkas. Reasoning, nonmonotonicity and learning in connectionist networks that capture propositional knowledge. *Artificial Intelligence*, 77(2):203–247, 1995.
- 59 John L. Pollock. Self-defeating arguments. *Minds and Machines*, 1:367–392, 1991.
- 60 John L. Pollock. A theory of defeasible reasoning. *International Journal of Intelligent Systems*, 6(1):33–54, 1991.
- 61 Nico Potyka. Extending modular semantics for bipolar weighted argumentation. In *Proceedings of the Eighteenth International Conference on Autonomous Agents and MultiAgent Systems, AAMAS'19*, pages 1722–1730, 2019.
- 62 Henry Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law, ICAL'05*, pages 85–94, 2005.
- 63 Henry Prakken. Modelling accrual of arguments in aspic+. In *In Seventeenth International Conference on Artificial Intelligence and Law (ICAIL'19)*, pages 285–297, 2019.

- 64 Antonio Rago and Francesca Toni. Quantitative argumentation debates with votes for opinion polling. In *Proceedings of the Twentieth International Conference on Principles and Practice of Multi-Agent Systems, PRIMA'17*, pages 369–385, 2017.
- 65 Yakoub Salhi. On an argument-centric persuasion framework. In *Proceedings of the Eighteenth International Conference on Autonomous Agents and MultiAgent Systems, AAMAS'19*, pages 1279–1287, 2019.
- 66 Bart Verheij. Accrual of arguments in defeasible argumentation. In *Proceedings of the Second Dutch/German Workshop on Nonmonotonic Reasoning*, pages 217–224, 1995.
- 67 Serena Villata, Guido Boella, and Leendert van der Torre. Argumentation patterns. In *Proceedings of the Eighth International Workshop on Argumentation in Multi-Agent Systems (ArgMAS'11)*, pages 133–150, 2011.