Arguing about the existence of conflicts [1]

Giuseppe PISANO [a,2], Roberta CALEGARI [a], Henry PRAKKEN [b,c], and Giovanni SARTOR [a]

[a] *Alma AI—Alma Mater Research Institute for Human-Centered Artificial Intelligence,* ALMA MATER STUDIORUM–*Università di Bologna, Italy*

[b] *Department of Information and Computing Sciences, Utrecht University, the Netherlands*

[c] *Faculty of Law, University of Groningen, the Netherlands*

Abstract. In this paper we formalise a meta-argumentation framework as an ASPIC+ extension which enables reasoning about conflicts between formulae of the argumentation language. The result is a standard abstract argumentation framework that can be evaluated via grounded semantics.

Keywords. meta-argumentation, argumentation, ASPIC+

1. Introduction

Meta-arguments support conclusions about other arguments, their interaction, their composition or their evaluation. For instance, a meta-argument may conclude that other arguments are in conflict or that one of them is preferred over the other, or it may provide new rules or facts that can be used in building arguments.

Meta-argumentation has received little attention thus far. As discussed in [1] there are various approaches to generate argumentation frameworks (AFs) in terms of accounts of the structure of arguments and their relations (e.g. ASPIC+, ABA, classical argumentation, DeLP). However, most of these approaches regard rule sets, specifications of conflicts and preferences as given. In the reality of adversarial debate, these things can also be argued about. Hence the importance of meta-argumentation.

In this paper we shall focus on a specific application of meta-argumentation to the conflict function of an argumentation theory, namely, assessing whether there is a conflict between two propositions in the argumentation language, i.e., whether the arguments concerning those propositions are incompatible so that accepting one of them entails rejecting the other.

Example 1 (Gender example) *To ground the discussion on a practical example, let us consider a legal example concerning a case of gender identity. Let us consider the case of Sue. She wants to compete in the woman's chess championship but the organisers argue*

[1] Research partially supported by the "CompuLaw" project, funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 833647).

[2] Corresponding Author: Giuseppe Pisano; E-mail: g.pisano@unibo.it.

that this would be impossible because legally she has been assigned the male gender, as proven by her passport. However, Sue is bigender – i.e. she identifies as both male and female simultaneously – and thinks that she should have the right to compete in the championship. To decide the case we should first decide on the existence of a conflict between the concepts of man and woman: are they in conflict – gender binarism discards her claim of being both man and woman at the same time –, or can the two concepts coexist according to the principle of self-determination? To encode the case at hand, the argumentation model should allow conflicts to be formalised, i.e., a meta-argumentation model is required.

It should now be more clear how the ability to include conflicts in the arguable content of a theory is fundamental in cases like the one we described above. The point of this work is whether this can be done while maintaining the compatibility with traditional argumentation methods and models—namely, Dung's semantics [2]. In this paper we focus on grounded semantics. For one reason, grounded semantics allows efficient use of the model in a real computational scenario—grounded semantics is the only one having polynomial complexity. Moreover, the use of grounded semantics only, allowed the authors – and hopefully the readers – to better focus on the fundamental ideas and mechanisms behind the proposed model, without the need to deal with the complexity of other semantics. For these reasons, the extension to other semantics is left to future work.

The main idea behind this work is to start from a standard structured argumentation framework – like ASPIC+ [3] – and expand its definitions to deal with meta reasoning over conflicts. We address meta-argumentation by using the mechanism presented in [4] for preferences and adapting it to conflicts, i.e., in representing attacks and conflicts through arguments, which in their turn, may be subject to attack. In this way, we can model meta-argumentation while preserving the semantics of standard abstract argumentation [2].

The paper is organised as follows. Background notions are discussed in Section 2, while Section 3 introduces the meta argumentation framework. Section 4 presents the related work and conclusions are drawn in Section 5.

2. Abstract Argumentation and Argumentation Theories

In this section we introduce the standard definitions for argumentation frameworks based on Dung's semantics [2] and for ASPIC+.

Definition 1 (Argumentation framework) *An argumentation framework AF is a tuple $<A, \rightsquigarrow>$, where A is a set of arguments and $\rightsquigarrow$ is a binary relation (attack relation) over $A \times A$. We write $X \rightsquigarrow Y$ for $(X,Y) \in \rightsquigarrow$.*

The semantics for an argumentation framework is defined as follows.

Definition 2 (Semantics) *Let $< \mathscr{A}, \rightsquigarrow>$ be an AF and $S \subseteq \mathscr{A}$. S is conflict free iff there are no $A, B \in S$ such that $A \rightsquigarrow B$. For any $X \in \mathscr{A}$, X is acceptable with respect to $S \subseteq \mathscr{A}$ iff $\forall Y \in \mathscr{A}$, $Y \rightsquigarrow X$ implies that $\exists Z \in S$ s.t. $Z \rightsquigarrow Y$. Then:*

- *S is an admissible extension iff $X \in S$ implies that X is acceptable w.r.t. S;*

- *S is a complete extension iff $X \in S$ iff X is acceptable w.r.t. S;*
- *S is the grounded extension iff S is the set-inclusion minimal complete extension.*

Definition 3 (Argumentation system) *An argumentation system is a quadruple AS=$< L, R, n, \rhd >$ where:*

- *L is a logical language;*
- *$R = R_s \cup R_d$ is a set of rules. R_d is a set of defeasible rules in the form $\phi_0, ..., \phi_n \Rightarrow \phi$, R_s is a set of strict rules in the form $\phi_0, \dots, \phi_n \rightarrow \phi$, where $\phi_0, \dots, \phi_n, \phi$ are well-formed formulae in the L language;*
- *n is a naming function of the form $n : R \mapsto L$*
- *$\rhd$ is a non-symmetrical conflict relation over $L \times L$. We write $\phi \rhd \psi$ for $(\phi, \psi) \in \rhd$.*

Definition 4 (Knowledge base) *A knowledge base for an AS=$< L, R, n, \rhd >$ is a set $K \subseteq L$ consisting of two disjoint subsets K_s (the axioms) and K_p (the ordinary premises).*

Definition 5 (Argumentation theory) *An argumentation theory is a tuple AT=$< AS, K >$ where AS is an argumentation system and K is a knowledge base in AS.*

Given an argumentation theory, by chaining rules from the theory we can construct arguments, as specified in the following definition; cf. [5,6,7].

Definition 6 (Argument) *Starting from an argumentation theory AT=$< AS, K >$, an argument A is any structure obtained by applying the following steps a finite number of times*

1. *ϕ if $\phi \in K$ with: Prem(A)=$\{\phi\}$; Conc(A)=ϕ; Sub(A)=$\{\phi\}$; DefRules(A)=$\emptyset$; TopRule(A)=undefined.*
2. *$A_1, \dots, A_n \Rightarrow \psi$ if $A_1, \dots, A_n$ are arguments s.t. $\exists$ a rule $r = Conc(A_1), \dots, Conc(A_n) \Rightarrow \psi \in R_d$.*
   - *Prem(A)=$Prem(A_1) \cup \dots \cup Prem(A_n)$,*
   - *Conc(A)=ψ,*
   - *Sub(A)=$Sub(A_1) \cup \dots \cup Sub(A_n) \cup \{A\}$,*
   - *TopRule(A)=r,*
   - *DefRules(A)=$DefRules(A_1) \cup \dots \cup DefRules(A_n) \cup \{r\}$*
3. *$A_1, \dots, A_n \rightarrow \psi$ if $A_1, \dots, A_n$ are arguments s.t. $\exists$ a rule $r = Conc(A_1), \dots, Conc(A_n) \rightarrow \psi \in R_s$.*
   - *Prem(A)=$Prem(A_1) \cup \dots \cup Prem(A_n)$,*
   - *Conc(A)=ψ,*
   - *Sub(A)=$Sub(A_1) \cup \dots \cup Sub(A_n) \cup \{A\}$,*
   - *TopRule(A)=r,*
   - *DefRules(A)=$DefRules(A_1) \cup \dots \cup DefRules(A_n)$*

Given an argument A we write:

- $Prem(A)$, for the set of premises from K used in the argument;
- $Conc(A)$, for the conclusion of the argument;
- $Sub(A)$, for the set of subarguments of A;
- $DefRules(A)$, for the set of rules in R_d used to build the argument;

- $TopRule(A)$, for the rule from R used in A's last inference step.

The first condition deals with arguments generated using the knowledge base K. Using the second and third ones we can recursively apply rules from R on the generated arguments to generate new arguments.

We can produce attacks starting from arguments using the notion of conflict for an argumentation language L:

Definition 7 (Direct attack) *An argument A directly attacks an argument B iff A directly undercuts, directly undermines or directly rebuts B where:*

- *A directly undercuts B iff* $Conc(A) \rhd n(TopRule(B))$ *and* $TopRule(B) \in R_d$*;*
- *A directly rebuts argument B iff* $Conc(A) \rhd Conc(B)$ *and* $TopRule(B) \in R_d$*;*
- *A directly undermines argument B iff* $B \in K_p$ *and* $Conc(A) \rhd B$

Definition 8 (Attack) *We say that argument A attacks argument B if A directly attacks* $B' \in Sub(B)$*.*

Then we can build an abstract argumentation framework as:

Definition 9 (Abstract argumentation framework) *Let* AT *be an argumentation theory* $< AS, K >$*. An abstract argumentation framework defined by* AT*, is a tuple* $< \mathscr{A}, \rightsquigarrow >$ *where:*

- $\mathscr{A}$ *is the set of all arguments constructed from* AT *according to Definition 6;*
- *for any arguments* X *and* $Y \in \mathscr{A}$, $X \rightsquigarrow Y$ *iff* X *attacks* Y

In the following sections, we will extend this model to base the $\rhd$ relation and consequently the $\rightsquigarrow$ relation on the content of the argumentation theory—i.e., shape the applicable conflicts inside the framework that we are evaluating.

The desiderata as a result is a standard abstract argumentation framework, thus preserving the possibility to evaluate it through the semantics given in Definition 2.

3. Reasoning with conflicts

According to Definition 3, the conflict relation is a fixed part of the argumentation system and attacks between arguments are determined by conflicts between the conclusion of the attacking argument and a premise, rule name, or conclusion of the directly attacked argument. The main idea underpinning the extension for dealing with meta-argumentation is to make the conflict relation dynamic, allowing arguments to argue for or against the existence of conflicts. In such a way we define an abstract argumentation framework that – once evaluated according to a standard Dung's semantics – produces admissible extensions containing both the arguments arguing on conflicts and arguments whose admissibility is influenced by these conflicts.

Let us start providing definitions for an argumentation language L enabling conflicts between elements of L to be stated, i.e., enabling reasoning with conflicts. We do that by introducing in the language a binary predicate – *conf* – putting in relation arbitrary formulae from the language itself. The introduced predicate will provide a way to express the content of the conflict relation $\rhd$ and use it in the argumentation process. Furthermore, we introduces wff's att(A) for any argument constructible with any possible set of rules over the new language.

Definition 10 (Conflict-based argumentation language) *Given an argumentation language L we define a argumentation language for reasoning with conflicts L_c as the smallest argumentation language $L_c = L \cup \{conf(\psi,\phi) | \psi,\phi \in L_c\} \cup \{attA \mid A$ is constructible with any set of rules over $L_c\}$.*

Now let us consider L_c a language as in Definition 10. We can build an argumentation system AS=$< L_c, R, n, \emptyset >$ and consequently an argumentation theory AT=$< AS, K >$, and use them to build an abstract argumentation framework AF=$< \mathscr{A}, \rightsquigarrow >$ using Definition 9. Note that, since $\triangleright = \emptyset$, the attack set $\rightsquigarrow$ in AF will be empty as well.

Now, let us extend the AF framework so defined to introduce attacks derived from the conflicts reified in the L_c language. In such a way the status of an attack is bound to the status of the argument claiming the conflict that generated it.

First, let us define an argument for each potential attack deriving from $conf$ predicates. Accordingly, attacks could be evaluated w.r.t. the semantics applied to the framework.

Definition 11 (Conflict-based direct attack argument) *A conflict-based direct attack argument X stating that argument W, based on conflict argument W', attacks argument Z, has the form $W, W' \Rightarrow att(Z)$ where:*

- *$Conc(W) = \phi$, $Conc(W') = conf(\phi,\psi)$ and*
  - ∗ *$n(TopRule(Z)) = \psi$ and $TopRule(Z) \in R_d$, or*
  - ∗ *$Conc(Z) = \psi$ and $TopRule(Z) \in R_d$ or*
  - ∗ *$Conc(Z) = \psi$ and $Z \in K_p$*

We define:

- *$Conc(X) = att(Z)$*
- *$Sub(X) = Sub(W) \cup Sub(W') \cup \{X\}$*

Let us write $DirectAttack(X)$ to indicate that X is a direct attack argument.

Thus to construct a *direct attack argument* $W, W' \Rightarrow att(Z)$ against Z it must be the case that two arguments are available, argument W, and argument W', the latter claiming that the conclusion of W is in conflict with the relevant element of Z (i.e., the name of Z's top rule or Z's conclusion). The status of the direct attack arguments will depend on the status of both W and W'.

We leverage direct attack arguments to build the actual *attack* set of the meta argumentation framework.

Definition 12 (Conflict-based attack) *A direct attack argument $W, W' \Rightarrow att(Z)$ attacks any argument Z' such that $Z \in Sub(Z')$.*

Thus, a direct attack argument $W, W' \Rightarrow att(Z)$ attacks not only its direct target Z, but also any argument Z' of which Z is a subargument. The success of the attack will depend not only on the status of W, but also on the status of W' which asserts that W and Z are in conflict.

These elements are merged together in a *Conflict-based Argumentation Framework*.

Definition 13 (Conflict-based Argumentation Framework) *Given an argumentation theory AT=$<< L_c, R, n, \emptyset >, K >$ with L_c being a conflict-based argumentation language, the conflict-based argumentation framework of AT is the tuple $< \mathscr{A}_1 \cup \mathscr{A}_2, \rightsquigarrow >$ where:*

- *$\mathscr{A}_1$ is the set of all arguments constructed from AT according to Definition 6;*
- *$\mathscr{A}_2$ is the set of all direct attack arguments constructed from AT and A_1 according to definition 11;*
- *$X \rightsquigarrow Z$ iff X attacks Z according to definition 12*

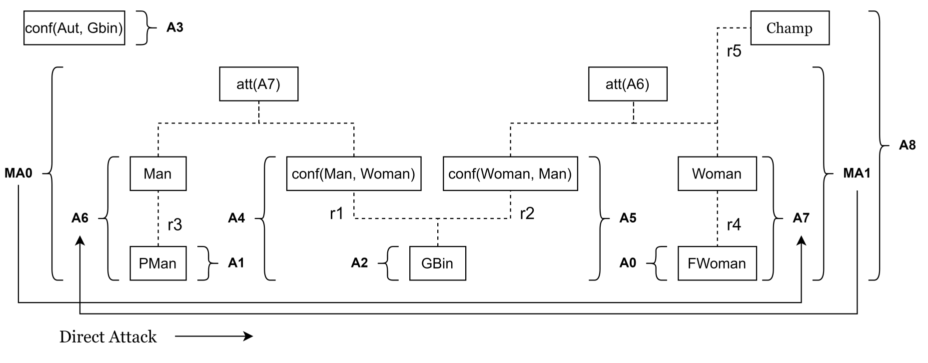
Conflict freeness, acceptability, admissible, complete, grounded extension are defined as in Definition 2.

The set $\mathscr{A}_2$ contains all attack arguments that can be generated by using the arguments in A_1, according to definition 13. For an attack argument $W, W' \Rightarrow att(Z)$ to be established according to an argumentation semantics, it is necessary that also W' is acceptable, i.e., that it is established that an acceptable conflict between W and Z exists. Only in this case W will bring an attack against Z.

Let us now provide some examples for our framework, in accordance the legal example introduced in Example 1. We use the following abbreviations:

- Champ = Sue can compete in the women's chess championship
- FWoman = Sue is bigender and identifies herself also as a woman
- PMan = Sue's passport identifies her as a man
- Aut = Every person has the right to self-determine their gender
- GBin = Every person's gender is determined by their birth sex, either male or female

Example 2 (Gender example: formalization) *Let us consider the theory where $R_d = \{$ r1 : Gbin $\Rightarrow conf(Man, Woman)$; r2 : Gbin $\Rightarrow conf(Woman, Man)$; r3 : PMan $\Rightarrow$ Man; r4 : FWoman $\Rightarrow$ Woman; r5 : Woman $\Rightarrow$ Champ $\}$ with the following facts $K_p = \{$ FWoman, PMan, GBin, $conf(Aut, Gbin)$ $\}$, $K_s = \emptyset$, $R_s = \emptyset$. Accordingly to the above definitions, we can then build the following arguments:*



The attacks are $MA0 \rightsquigarrow A7$, $MA0 \rightsquigarrow A8$, $MA1 \rightsquigarrow A6$, $MA0 \rightsquigarrow MA1$, $MA1 \rightsquigarrow MA0$. If we apply Dung's grounded semantics to the framework we obtain the extension $\{A0, A1, A2, A3, A4, A5\}$—i.e. the incompatibility between Sue's official gender and her other perceived identity ($A4, A5$) prevents her to compete in the championship.

Sue is not happy with the final decision and decides to appeal claiming that her right to self-determination has not been taken into due consideration. The case is evaluated again with the new information: $K'_p = \{Aut\} \cup K_p$. Two new arguments are obtained:

A9 : Aut
MA2 : A9, A3 ⇒ att(A2)

The new attacks are $MA2 \rightsquigarrow A2$, $MA2 \rightsquigarrow A4$, $MA2 \rightsquigarrow A5$, $MA2 \rightsquigarrow MA0$, $MA2 \rightsquigarrow MA1$. Applying again Dung's grounded semantics to the framework we obtain the extension $\{A0,A1,A3,A6,A7,A8,A9,MA2\}$—i.e. the problem on Sue's identity is resolved discarding the binary view on genders (A2, A4, A5), according to the principle of self-determination (A9). Consequently, Sue's perceived genders are both present in the extension, and she is free to compete in the championship. Indeed, the CAF was able to integrate the new knowledge and use it to revise the status of the propositional conflicts in the argumentation theory as expected. Both the original argumentation graph and the revised one are presented in Figure 1.

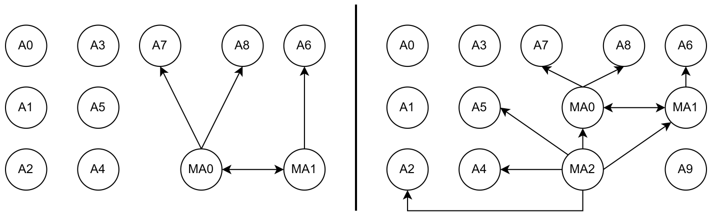


Figure 1. Conflict-based Argumentation frameworks from Example 2

3.1. Properties

We now proceed to demonstrate two important properties of the constructed framework. Intuitively, we would expect that a conflict that has been proven to exist at the meta-level – i.e. via the conflict-based framework –, indeed exists at the object level, leading to the same set of attacks and, consecutively, to the same extension. In other words, what is true according to the conflict-based framework should remain true when the verified conflicts are applied a priori as in the original ASPIC$^+$ model. To demonstrate this important property, let us introduce the notion of an *Equivalent Standard AF*.

To start, we define a way to construct a standard argumentation framework on the basis of a conflict-based argumentation framework. The basic idea is that starting with a conflict-based argumentation framework and an extension of it, we construct a standard argumentation framework having a corresponding extension according to the same semantics.

Let us consider a conflict-based argumentation framework CAF=$< \mathscr{A}, \rightsquigarrow >$ and one of its extensions E. To construct the equivalent argumentation framework EAF, we first remove from $\mathscr{A}$ (a) all attack arguments that are supported by those conflict arguments that are in the extension, and (b) all attack arguments that are supported by a conflicting argument that is attacked by the extension. Only attack arguments that are neither included in E nor attacked by it are left in the EAF's arguments set. Accordingly, the EAF's attack relation is constructed using those conflicts claimed by the arguments in E.

Definition 14 (Equivalent Standard AF) *Given a conflict based argumentation framework CAF=$< \mathscr{A}, \rightsquigarrow >$ having an extension E according to semantics σ, we define an equivalent standard argumentation framework EAF=$< \mathscr{A}', \rightsquigarrow' >$ where:*

1. $\mathscr{A}' = \mathscr{A} \setminus B \cup C$ *where:*
   (a) $B = \{a \in \mathscr{A} \mid \exists b \in E$ *such that* $Conc(b) = conf(\phi, \psi)$ *and a is a direct attack argument of the form* $W, b \Rightarrow Z\}$*;*
   (b) $C = \{a \in \mathscr{A} \mid \exists b \in \mathscr{A}$ *such that* $Conc(b) = conf(\phi, \psi)$ *and b is attacked by E and a is a direct attack argument of the form* $W, b \Rightarrow Z\}$.
2. $\rightsquigarrow' = \rightsquigarrow_{|\mathscr{A}' \times \mathscr{A}'} \cup \{(a,b) \mid a, b \in \mathscr{A}'$ *and* $\exists c \in E$ *such that* $Conc(c) = conf(\phi, \psi)$ *and* $\exists\, b' \in Sub(b)$ *s.t. a directly attacks* b' *(Definition 7) according to the conflict* $\phi \rhd \psi\}$.

Proposition 1 *Consider a finitary CAF=$< \mathscr{A}, \rightsquigarrow >$ and its corresponding EAF=$< \mathscr{A}', \rightsquigarrow' >$ built on the grounded extension E and having grounded extension E'. Then $E \cap \mathscr{A}' = E'$.*

Proof 1 *Let's consider an argumentation framework CAF $=< \mathscr{A}, \rightsquigarrow >$. We call the characteristic function of CAF the function $F : 2^{\mathscr{A}} \rightarrow 2^{\mathscr{A}}$ such that $F(Args) = \{X | \forall Y$ such that $Y \rightsquigarrow X$, then $\exists Z \in Args$ such that $Z \rightsquigarrow Y\}$ where $Args \subseteq \mathscr{A}$. Let us consider grounded extension as the minimal conflict-free fixed point of the characteristic function F—i.e. the union of a sequence $E_0, \ldots, E_n$ obtained by iterative application of the F function on the empty set, and where $E_0 = \emptyset$. We prove that $E \cap \mathscr{A}' = E'$.*

We first prove that $E \cap \mathscr{A}' \subseteq E'$. Suppose $a \in E \cap \mathscr{A}'$. We prove that $a \in E'$ as follows.

Base case: *a has no attackers in $\mathscr{A}$ according to $\rightsquigarrow$ so $a \in E_1 \cap \mathscr{A}'$. Then there can only be attackers of a in $\mathscr{A}'$ according to $\rightsquigarrow'$ if there is a relevant conflict argument b in E that says that the conclusion of some argument $x \in \mathscr{A}$ conflicts with a's conclusion. But then there exists a direct attack argument $x, b \Rightarrow att(a)$ in $\mathscr{A}$, which contradicts that a has no attackers in $\mathscr{A}$. So $x \notin \mathscr{A}'$, so a has no attackers in $\mathscr{A}'$, so $a \in E'$.*

Induction step: *Assume that all arguments in $E_{i-1} \cap \mathscr{A}'$ are in E'. Consider any $a \in E_i$. Any $b \in \mathscr{A}'$ such that $b \rightsquigarrow' a$ is such that $b \rightsquigarrow a$ or $b \not\rightsquigarrow a$. First, any such b such that $b \rightsquigarrow a$ is attacked by E_{i-1} according to $\rightsquigarrow$. Then by the induction hypothesis, if $b \in \mathscr{A}'$, then b is also attacked by E'. Next, consider any such b such that $b \not\rightsquigarrow a$. Then there is a direct attack argument $m \in \mathscr{A}$ of the form $b, X' \Rightarrow att(a')$ with $a' \in Sub(a)$. Then $m \rightsquigarrow a$ so there exists an $m' \in E_{i-1}$ such that $m' \rightsquigarrow m$. Note that m is a direct attack argument, so m is of the form $c, Z \Rightarrow att(b')$ with $b' \in Sub(b)$. By closure of E under subarguments (an easy adaptation of the same result on standard ASPIC+), c and Z are also in E_{i-1}. But then by the induction hypothesis $c \in E'$ and $c \rightsquigarrow' b$. So $a \in E'$.*

We next prove that $E' \subseteq E \cap \mathscr{A}'$. Suppose $a \in E'$. We prove that $a \in E$ as follows.

Base case: *a has no attackers in $\mathscr{A}'$ so $a \in E'_1$. Consider any $b \in \mathscr{A}$ such that $b \rightsquigarrow a$. Then b is a direct attack argument of the form $c, X \Rightarrow att(a)$, with X a conflict argument that says that the conclusion of argument $c \in \mathscr{A}$ conflicts with a's conclusion. But since a has no attackers in $\mathscr{A}'$, we have that $b \notin \mathscr{A}'$ because of either condition (1a) or condition (1b) of Definition 14. In the case of (1b), b is attacked according to $\rightsquigarrow$ on X by an argument in E. In the case of (1a), we have $X \in E$, so, according to condition (2) of Definition 14, we have that $c \rightsquigarrow' a$. But this contradicts that a has no attackers in $\mathscr{A}'$. The two cases together prove that $a \in E$.*

Induction step: *Assume that all arguments in E'_{i-1} are in E. Consider any $a \in E'_i$. Then all $b \in \mathscr{A}'$ such that $b \rightsquigarrow' a$ are attacked by E'_{i-1} according to $\rightsquigarrow'$. Then b could be either a regular argument or a direct attack argument of the form $c, X \Rightarrow att(a')$ with $a' \in Sub(a)$. In the latter case, since $b \in \mathscr{A}$ then, by the induction hypothesis, b is also attacked by E according to $\rightsquigarrow$. In the first case, since $b \rightsquigarrow' a$, there must exist a direct attack argument $m \in \mathscr{A}$ of the form $b, X \Rightarrow att(a')$ with $a' \in Sub(a)$ such that $m \rightsquigarrow a$. But, by induction hypotheses, b and m are both attacked according to $\rightsquigarrow$ by E. The two cases together prove that $a \in E$.*

Since $E \cap \mathscr{A}' \subseteq E'$ and $E' \subseteq E \cap \mathscr{A}'$ then $E' = E \cap \mathscr{A}'$.

The second property we want to demonstrate builds on top of what we have just proven. We have seen that it is possible to move the conflicts in the grounded extension at the object level without altering the results. However, the resulting *Equivalent Standard AF* still contains the meta-level attack arguments that are not in the extension or attacked by a member of it. The question is whether there are cases in which the conflict framework can be completely transformed into a regular argumentation framework. The implication of this finding would be straightforward: the Conflict-based framework would be a generalisation of a regular abstract argumentation framework. This is a fundamental property for every model trying to provide a conservative extension like ours.

The next proposition shows that a *Conflict-based Argumentation Framework* is a generalisation of a standard abstract argumentation framework.

Proposition 2 *Consider a CAF=$< \mathscr{A}, \rightsquigarrow >$ and its corresponding equivalent EAF=$< \mathscr{A}', \rightsquigarrow' >$ built on the σ extension E. If $\forall x \in \{a \in \mathscr{A} | Conc(a) = conf(\phi, \psi)\}$ we have that either $x \in E$ or $\exists (d, x) \in \rightsquigarrow$ s.t. $d \in E$, then EAF is a regular argumentation framework as in Definition 9.*

Proof 2 *By Definition 14 if all the conflict arguments are either in the extension or attacked by a member of it, then all the Direct Attack Arguments can be discarded leaving only the arguments produced using Definition 6. The attack set would then be given by the set of conflicts claimed by the argument in the extension using Definition 7. Consequently, the result is a regular argumentation framework as in Definition 9.*

Example 3 (Gender Example: propositions) *To ground Proposition 1 and 2 let us consider again the framework in Example 2.First, we have to build the Equivalent Standard AF using the results of Dung's grounded semantics. All the conflict arguments are either in the extension ($A3$) or attacked by a member of the extension ($A4$, $A5$). Accordingly, we can delete from the set of arguments the linked attack arguments ($MA0$, $MA1$, $MA2$), and use the conflict claimed by $A3$ to build the new attack set. We have only three attacks, $A9 \rightsquigarrow A2$, $A9 \rightsquigarrow A4$ and $A9 \rightsquigarrow A5$, as shown in Figure 2.*

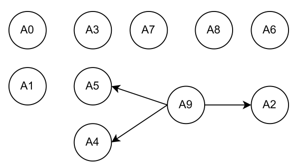


Figure 2. Equivalent Standard AF from Example 3

If we apply the grounded semantics to the framework then we obtain the extension $\{A0, A1, A3, A6, A7, A8, A9\}$*—the same as in the Conflict-based AF but without attack arguments, as claimed by Property 1. It is worth noting that we obtained a regular framework as result: all the conflicts are known a priori and the same framework could have been built using the standard ASPIC*$^{+}$ *definitions.*

In the general case, however, we cannot have a complete equivalency between a CAF and a regular framework. Indeed, if an argument for $conf(\psi, \phi)$ is undecided – neither in the extension nor attacked by one of its members –, then the uncertainty can be propagated to the attack argument and then to the attacked argument, thus preventing them to be part of the extension. We could not obtain the same result without considering the Direct Attack Argument, because the absence of the conflict would potentially allow the attacked argument to be accepted without considering the potential uncertainty in the state of the conflict. In other words, a CAF framework is capable of conveying more information on the state of an attack w.r.t. a standard argumentation framework, thus making the transformation to a regular framework impossible in the general case.

Example 4 (Partial Transformation) *Let us consider the theory where* $K_p = \{p, q, r, -r\}$ *and* $K_s = \{conf(r, -r), conf(-r, r)\}$ *and* $R_d = \{r => conf(p, q)\}$*. Starting from this theory we can build the Conflict-based framework and then the Equivalent one as shown in Figure 3.*

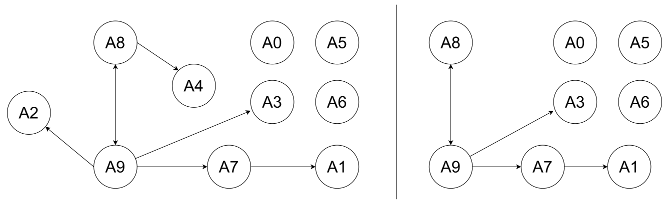
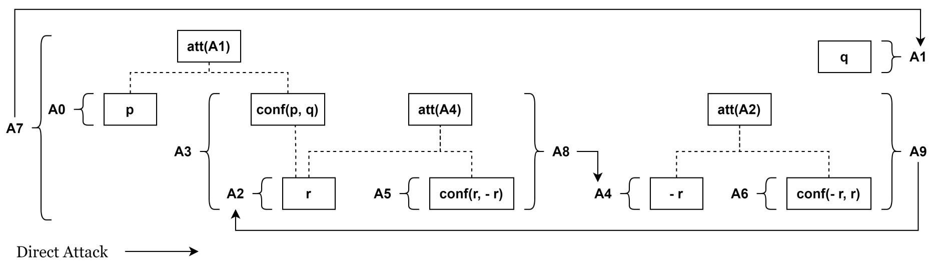


Figure 3. Conflict-based Argumentation framework from Example 4 on the left, Equivalent framework on the right.



If we apply Dung's grounded semantics to the frameworks, in both cases we obtain the extension $\{A0, A5, A6\}$*. It can be noticed that the Equivalent framework still contains an attack argument (*$A7$*) due to the uncertainty in* $A3$*'s evaluation. Indeed, without knowing if* $A3$ *is in the extension or definitely rejected – i.e. attacked by a member of the extension –, it is impossible to decide whether* $A0$ *should attack* $A1$ *or not in the Equivalent Standard AF. Consequently, every alteration of the Equivalent attack set on the basis of this conflict would lead to a possible modification in the semantics results—i.e. the attack argument* $A7$ *with the connected attacks must be preserved in the Equivalent AF.*

4. Related Research

Modgil & Bench-Capon [8] introduce the notion of meta-level argumentation frameworks. The arguments of meta-level argumentation frameworks make claims about object-level abstract argumentation frameworks according to the theory of such frameworks, for example, "A is in a preferred extension of AF" or "argument A in AF defeats argument B in AF". Constraints are formulated on the attacks of the meta-level framework to ensure that such statements are correct with respect to the object level. For example, "y defeats x" attacks "x is justified". This allows the formalisation of Dung's theory of abstract argumentation frameworks in meta-level argumentation frameworks that have the same semantics as Dung's original frameworks. Moreover, Modgil & Bench-Capon show that the same approach can be used to formalise variants of Dung-style argumentation frameworks, such as preference- and value based AFs and extended AFs. In a similar way, Boella & al. [9] develop a general methodology for instantiating Dung's original argumentation frameworks starting from extended argumentation frameworks through a flattening technique—comparably to what is done in [10]. The resulting framework operates on meta-arguments, for example in the form "argument A is accepted" while remaining in the formal framework of Dung's argumentation theory. While these approaches are theoretically very interesting, they do not specify the structure of arguments at the object level and therefore seem less suitable for knowledge representation.

Moving beyond abstract argumentation, [11] introduces a variant of defeasible logic, Defeasible Meta-Logic, to represent defeasible meta-theories, by proposing algorithms to compute the (meta-)extensions of such theories, and by proving their computational complexity.

Wooldridge & al. [12] develop a completely different approach for dealing with the meta-argumentative nature of argument systems. The work proposes a hierarchical first-order meta-logic, producing a three tiers argument system. Level 0 contains statements on the object domain, level 1 introduces the notion of arguments and acceptability, while level 2 is used to reason on the structure of arguments and their relations. This formalism – because of the required hierarchical representation –, although enabling a clear separation between meta- and object- level concepts, could result in decreased flexibility in the formalisation of the knowledge in the system.

A limited kind of meta-argumentation can be found in argumentation frameworks that allow for arguments about preferences. In [13] conflicts between mutually rebutting arguments are decided by preferences, which are established by arguments included in the same argumentation framework. A fix-point semantics is used to compute extensions including preference arguments. Reasoning about preferences has been recently modelled by introducing a new kind of attack, namely, a preference-based attack against attacks [14]. Dung & al. [4] expands this idea, by having a framework that includes attack arguments, as well as preference attack arguments against attacks. In this way, the framework obtained can be evaluated by using standard Dung semantics.

5. Conclusions

Our paper has presented a meta-argumentation framework for reasoning over conflicts. In particular, we have provided an ASPIC$^+$ extension allowing the encoding of conflicts between formulae in the argumentation language. The conflicts – on which can be argued in the framework – are exploited to build meta-arguments representing attacks between arguments. The result is a framework in which the set of valid attacks is dynamically connected to the acceptability status of the conflicts used to derive them. In this way, we have modelled meta-argumentation while preserving the semantics of standard abstract argumentation introduced by [2].

At the moment, this work is limited to grounded semantics only. A natural direction for a future extension is to also provide formal proofs of the framework soundness for other Dung's semantics—i.e. complete, stable, preferred. Future work will also be devoted to comparing with alternative approaches, e.g. [12] as applied in [15], and extending the model so as to include other meta-components in the framework—e.g. conditional preferences [4] and nested meta-rules [11].

References

[1] Besnard P, Garcia A, Hunter A, Modgil S, Prakken H, Simari G, et al. Introduction to structured argumentation. Argument & Computation. 2014;5(1):1-4.

[2] Dung PM. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. Artificial Intelligence. 1995;77(2):321-58.

[3] Prakken H. An abstract framework for argumentation with structured arguments. Argument and Computation. 2010;1(2):93-124.

[4] Dung PM, Thang PM, Son TC. On Structured Argumentation with Conditional Preferences. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019. Honolulu, Hawaii, USA: AAAI Press; 2019. p. 2792-800.

[5] Caminada M, Amgoud L. On the Evaluation of Argumentation Formalisms. Artificial Intelligence. 2007;171(5—6):286-310.

[6] Modgil S, Prakken H. The $ASPIC^{+}$ framework for structured argumentation: a tutorial. Argument & Computation. 2014;5(1):31-62.

[7] Vreeswijk G. Abstract Argumentation Systems. Artif Intell. 1997;90(1-2):225-79. Available from: `https://doi.org/10.1016/S0004-3702(96)00041-0`.

[8] Modgil S, Bench-Capon TJM. Metalevel Argumentation. Journal of Logic and Computation. 2011;21:959-1003.

[9] Boella G, Gabbay DM, van der Torre LWN, Villata S. Meta-Argumentation Modelling I: Methodology and Techniques. Studia Logica. 2009;93(2-3):297-355.

[10] Gabbay DM. Semantics for Higher Level Attacks in Extended Argumentation Frames Part 1: Overview. Studia Logica. 2009;93(2-3):357-81.

[11] Olivieri F, Governatori G, Cristani M, Sattar A. Computing Defeasible Meta-logic. In: Logics in Artificial Intelligence. Cham: Springer International Publishing; 2021. p. 69-84.

[12] Wooldridge M, McBurney P, Parsons S. On the Meta-Logic of Arguments. In: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems. New York, NY, USA: Association for Computing Machinery; 2005. p. 560-7.

[13] Prakken H, Sartor G. A dialectical model of assessing conflicting arguments in legal reasoning. Artificial Intelligence and Law. 1996;4:331-68.

[14] Modgil S, Prakken H. Reasoning about Preferences in Structured Extended Argumentation Frameworks. In: Baroni P, Cerutti F, Giacomin M, Simari GR, editors. Computational Models of Argument. Proceedings of COMMA 2010. IOS; 2010. p. 347-58.

[15] van der Weide TL, Dignum F. Reasoning about and Discussing Preferences between Arguments. In: Argumentation in Multi-Agent Systems - 8th International Workshop, ArgMAS 2011, Taipei, Taiwan, May 3, 2011, Revised Selected Papers. vol. 7543 of Lecture Notes in Computer Science. Springer; 2011. p. 117-35.