Modelling Support Relations between Arguments in Debates

Henry Prakken[1,2]

[1] Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
[2] Faculty of Law, University of Groningen, Groningen, The Netherlands

Abstract. Many formal modellings of structured argumentation presuppose a knowledge base from which arguments are constructed. However, in debate contexts there usually is no global knowledge base from which the debate participants construct their arguments. The question then arises how these formalisms can be used for evaluating debates. On issue here is how support relations between arguments put forward in a debate should be modelled. This paper develops a formal approach within the *ASPIC*$^+$ framework and compares it to approaches using bipolar abstract argumentation frameworks. It is argued that for a proper model of debate evaluation it is crucial to look at the structure of arguments, which casts doubt on the benefits of purely abstract models of debate evaluation.

1 Introduction

Imagine John Doe watching or reading a debate on a topic like 'Does global global warming exist?', 'Should the west bomb the IS?' or 'Should the Schengen area be terminated?'. After the debate is finished, John wants to determine which arguments and claims put forward in the debate are acceptable. He does not care about how the debate evolved over time or who said what, he just wants to look at the contents of the arguments. He wants to reconstruct how the various arguments support or attack each other, he wants to express whether he accepts their premises or inferences, he wants to express his preferences between conflicting arguments, and he may want to add some arguments of his own. This is what in this paper will be called 'evaluating a debate'. Evaluating a debate in this sense is largely a subjective matter, since different people can make different choices on all the points just mentioned. However, there are still rational constraints, namely, those formulated by informal and formal argumentation theory. In this paper the focus is on how the current formal models of argumentation constrain debate evaluation in the sense just explained. To this end, it will be assumed that a formalised version of the arguments put forward in a debate already exists. In practice the step from a natural-language debate to a formalised version is far from trivial but this step is not what this paper is about.

At first sight, there would seem to be no problem: if a formalised set of arguments in some argumentation logic already exists, then it would seem to suffice to simply apply the argumentation logic to the set of arguments. However, there is still a problem here, since many formal modellings of structured argumentation presuppose a knowledge base and a set of rules from which arguments have to be constructed (e.g. [22, 17, 9, 7, 12, 19, 13]). The problem is that in debate contexts there usually is no global knowledge base or agreed set of rules from which the debaters construct their arguments. One problem in particular is how to deal with arguments put forward by different participants of the debate and where the conclusion of one argument provides a premise of the other argument. Should the two arguments be regarded as a single complex argument on the basis of some global background assumed in the debate, where the inferential support relations inside the single combined argument captures the relation of support, or should the arguments be regarded as separate entities related through another notion of support? Cayrol & Lagasquie-Schiex argue in [3] for the latter solution on the ground that it would be less natural to regard the situation where one agent supports an argument stated by another agent as a case of revising the supported argument by the supporter. Instead it would be more natural to model both arguments as separate entities. This view on debate evaluation was one motivation for the development of so-called bipolar argumentation frameworks (BAFS), which add abstract support relations to the theory of abstract argumentation frameworks (AFs) originating from [5].

One aim of this paper is to argue that this criticism is not justified and that it is still better to model the support relation between such arguments by combining them into a single complex argument and capturing their support relation in the inferential support relations within the argument. To account for the fact that the two original arguments were possibly stated by different debate participants, these two arguments will not be ignored but will also be considered in the evaluation process. A second aim of this paper is to show how this evaluation process allows for a more subtle evaluation of arguments than in the theory of BAFs.

Since BAFs are defined on top of AFs, the present investigations will also assume the theory of AFs. A third aim of this paper then is to show that analysing support relations between arguments in debates cannot be done without an account of the structure of arguments and the nature of attacks. As such an account, the *ASPIC*$^+$ framework [19, 13–15] will be used, which allows a modelling of inferential support relations with its notion of a subargument. However, the main ideas of this paper also apply to formalisms for structured argumentation that do not precisely instantiate the *ASPIC*$^+$ framework.

This paper is organised as follows. After presenting the formal preliminaries in Section 2, a way to reconstruct the arguments put forward in a debate in *ASPIC*$^+$ will be proposed in Section 3. Then in Section 4 it will be shown that a combination of *ASPIC*$^+$ and a BAF approach with premise support for modelling debate evaluation has some disadvantages and that it is better to combine supporting and supported arguments in a single compound argument by using *ASPIC*$^+$'s subargument relation. This method will be formalised in Section 5, after which it will be shown in Section 6 how the method can be applied to evaluating debates. The paper will conclude in Section 7.

2 Formal preliminaries

In this section the formal frameworks used or discussed in this paper are reviewed.

An *abstract argumentation framework* (AF) is a pair $\langle \mathcal{A}, \mathcal{D} \rangle$, where $\mathcal{A}$ is a set of arguments and $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$ is a relation of defeat. The theory of AFs [5] identifies sets of arguments (called *extensions*) which are internally coherent and defend themselves against defeaters. An argument $A \in \mathcal{A}$ is *defended* by a set by $S \subseteq \mathcal{A}$ if for all $B \in \mathcal{A}$: if B defeats A, then some $C \in S$ defeats B. Then relative to a given AF, $E \subseteq \mathcal{A}$ is *admissible* if E is conflict-free and defends all its members; E is a *complete extension* if E is admissible and $A \in E$ iff A is defended by E; E is a *preferred extension* if E is a $\subseteq$-maximal admissible set; E is a *stable extension* if E is admissible and attacks all arguments outside it; and $E \subseteq \mathcal{A}$ is the *grounded extension* if E is the least fixpoint of operator F, where $F(S)$ returns all arguments defended by S. It holds that any preferred, stable or grounded extension is a complete extension. Finally, for $T \in \{$complete, preferred, grounded, stable$\}$, X is *sceptically* or *credulously* justified under the T semantics if X belongs to all, respectively at least one, T extension.

Several proposals exist for adding support relations to abstract argumentation frameworks, the best-known being [3]'s bipolar argumentation frameworks (BAFs). The literature on BAFs contains several proposals for definitions of conflict-freeness and admissibility. For now it is not necessary to commit to any specific proposal; therefore for now simply 'abstract argumentation frameworks with support (SuppAFs)' (a term borrowed from [21]) will be considered, which add a binary support relation to AFs. Thus SuppAFs are a triple $(\mathcal{A}, \mathcal{D}, \mathcal{S})$ where $\mathcal{D}$ and $\mathcal{S}$ are binary relations over a set $\mathcal{A}$. Depending on the definitions of conflict-freeness and admissibility and on which further constraints are added, SuppAFs may or may not be BAFs.

The ***ASPIC*$^+$ framework** [19, 13, 14] gives structure to Dung's arguments and defeat relation. It defines arguments as directed acyclic graphs formed by applying strict or defeasible inference rules to premises formulated in some logical language. Arguments can be attacked on their (non-axiom) premises and on their applications of defeasible inference rules. Some attacks succeed as *defeats*, as partly determined by preferences. The acceptability status of arguments is then defined by applying any of [5]'s semantics for abstract argumentation frameworks to the resulting set of arguments with its defeat relation. Since the initial paper [19] on *ASPIC*$^+$, several variants of the framework have been proposed. For present purposes, their differences do not matter. In this paper we will use the variant with symmetric negation and so-called 'defeat conflict-freeness' as presented in [15].

ASPIC$^+$ is not a system but a framework for specifying systems. It defines the notion of an abstract *argumentation system* as a structure consisting of a logical language $\mathcal{L}$ with a binary negation symbol $\neg$, a set $\mathcal{R}$ consisting of two subsets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules, and a naming convention n in $\mathcal{L}$ for defeasible rules in order to talk about the applicability of defeasible rules in $\mathcal{L}$. Informally, $n(r)$ is a wff in $\mathcal{L}$ which says that rule $r \in \mathcal{R}$ is applicable.

Definition 1. ***[Argumentation systems]*** *An* argumentation system *is a triple* $AS = (\mathcal{L}, \mathcal{R}, n)$ *where:*

- $\mathcal{L}$ *is a logical language with a binary negation symbol* $\neg$*.*
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ *is a finite set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form* $\{\varphi_1, \ldots, \varphi_n\} \rightarrow \varphi$ *and* $\{\varphi_1, \ldots, \varphi_n\} \Rightarrow \varphi$ *respectively (where* φ_i, φ *are meta-variables ranging over wff in* $\mathcal{L}$*), such that* $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$*.* $\varphi_1, \ldots, \varphi_n$ *are called the* antecedents *and* φ *the* consequent *of the rule.*[3]

[3] Below the brackets around the antecedents will be omitted.

– n *is a partial function from* $\mathcal{R}_d$ *to* $\mathcal{L}$*, which to rules in* $\mathcal{R}_d$*, a naming convention for defeasible rules.*

We write $\psi = -\varphi$ *just in case* $\psi = \neg\varphi$ *or* $\varphi = \neg\psi$*.*

Definition 2. ***[Knowledge bases]*** *A* knowledge base *in an* $AS = (\mathcal{L}, \mathcal{R}, n)$ *is a set* $\mathcal{K} \subseteq \mathcal{L}$ *consisting of two disjoint subsets* $\mathcal{K}_n$ *(the* axioms*) and* $\mathcal{K}_p$ *(the* ordinary premises*).*

Arguments can be constructed step-by-step from knowledge bases by chaining inference rules into directed acyclic graphs (which are trees if no premise is used more than once). In what follows, for a given argument the function `Prem` returns all its premises, `Conc` returns its conclusion, `Prop` and `Rules` return, respectively, all wff and all rules occurring in it, `Sub` returns all its sub-arguments and `TopRule` returns the last inference rule applied in the argument.

Definition 3. ***[Arguments]*** *An* argument A *on the basis of a knowledge base* $\mathcal{K}$ *over an argumentation system* AS *is any structure obtainable by applying one or more of the following steps finitely many times:*

1. φ *if* $\varphi \in \mathcal{K}$ *with:* $\texttt{Prem}(A) = \{\varphi\}$*;* $\texttt{Conc}(A) = \varphi$*;* $\texttt{Prop}(A) = \{\varphi\}$*;* $\texttt{Sub}(A) = \{\varphi\}$*;* $\texttt{TopRule}(A)$ = *undefined;* $\texttt{Rules}(A) = \emptyset$*.*
2. $A_1, \ldots, A_n \rightarrow/\Rightarrow \psi$ *if* $A_1, \ldots, A_n$ *are arguments such that there exists a strict/defeasible rule* $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi$ *in* $\mathcal{R}_s/\mathcal{R}_d$*.*
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$*;*
   $\texttt{Conc}(A) = \psi$*;*
   $\texttt{Prop}(A) = \texttt{Prop}(A_1) \cup \ldots \cup \texttt{Prop}(A_n) \cup \{\psi\}$*;*
   $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$*;*
   $\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi$*;*
   $\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow/\Rightarrow \psi\}$.

For any argument A *we define* $\texttt{Prem}_n(A) = \texttt{Prem}(A) \cap \mathcal{K}_n$ *and* $\texttt{Prem}_p(A) = \texttt{Prem}(A) \cap \mathcal{K}_p$*. Moreover, for any set* $\mathcal{A}$ *of arguments,* $\texttt{Prem}(\mathcal{A}) = \{\varphi \mid \varphi \in \texttt{Prem}(A)$ *for some* $A \in \mathcal{A}\}$*. The notations* $\texttt{Conc}(\mathcal{A})$*,* $\texttt{Prop}(\mathcal{A})$ *and* $\texttt{Rules}(\mathcal{A})$ *are defined likewise while* $\mathcal{R}_s(\mathcal{A}) = \texttt{Rules}(\mathcal{A}) \cap \mathcal{R}_s$ *and* $\mathcal{R}_d(\mathcal{A}) = \texttt{Rules}(\mathcal{A}) \cap \mathcal{R}_d$*.*

Arguments can be attacked in three ways: on their premises (undermining attack), on their conclusion (rebutting attack) or on an inference step (undercutting attack). The latter two are only possible on applications of defeasible inference rules.

Definition 4. ***[Attack]*** A attacks B *iff* A undercuts*,* rebuts *or* undermines B*, where:*
• A undercuts *argument* B *(on* B'*) iff* $\texttt{Conc}(A) = -n(r)$ *and* $B' \in \texttt{Sub}(B)$ *such that* B'*'s top rule* r *is defeasible.*
• A rebuts *argument* B *(on* B'*) iff* $\texttt{Conc}(A) = -\varphi$ *for some* $B' \in \texttt{Sub}(B)$ *of the form* $B''_1, \ldots, B''_n \Rightarrow \varphi$*.*
• *Argument* A undermines B *(on* B'*) iff* $\texttt{Conc}(A) = -\varphi$ *for some* $B' = \varphi$*,* $\varphi \notin \mathcal{K}_n$*.*

Argumentation systems plus knowledge bases form argumentation theories, which induce structured argumentation frameworks.

Definition 5. ***[Structured Argumentation Frameworks]*** *Let AT be an argumentation theory $(AS, \mathcal{K})$. A structured argumentation framework (*SAF*) defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ where $\mathcal{A}$ is the set of all arguments on the basis of $\mathcal{K}$ in AS, $\preceq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.*

The *ASPIC*$^+$ notion of *defeat* can then be defined as follows. Undercutting attacks succeed as *defeats* independently of preferences over arguments, since they express exceptions to defeasible inference rules. Rebutting and undermining attacks succeed only if the attacked argument is not stronger than the attacking argument ($A \prec B$ is defined as usual as $A \preceq B$ and $B \npreceq A$).

Definition 6. ***[Defeat]*** A defeats B *iff: A undercuts B, or; A rebuts/undermines B on B' and $A \nprec B'$.*

Abstract argumentation frameworks are then generated from SAFs as follows:

Definition 7 (Argumentation frameworks). *An* abstract argumentation framework (AF) corresponding to a $SAF = \langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ *is a pair $(\mathcal{A}, \mathcal{D})$ such that $\mathcal{D}$ is the defeat relation on $\mathcal{A}$ determined by SAF.*

Then several ways are possible of using the theory of AFs for evaluating conclusions of arguments. One is to say that a formula φ is a skeptical (credulous) consequence of a SAF iff an argument with conclusion φ is in all (some) extensions of the AF corresponding to SAF. Other definitions are possible; for present purposes their differences do not matter.

3 Formalising debates

Assume that in a debate the participants construct the arguments in their own internal *ASPIC*$^+$ argumentation theory. It cannot in general be assumed that these theories have the same language, rules and knowledge base. All an outsider can observe is the arguments, that is, their premises and inferences. The task then is to construct a global argumentation theory that generates these arguments and that does not contain any information that is not contained in these arguments. To make this well-defined, such an AT will be defined indirectly in the notion of a debate-generated SAF, otherwise the set of arguments that generates the AT cannot be easily stated in a well-defined way.

Let a partial SAF be defined as a SAF except that $\mathcal{A}$ is any set of arguments constructible on the basis of $\mathcal{K}$ in AS (so not all constructible arguments need to be in $\mathcal{A}$). Then:

Definition 8. *Let a* debate-generated SAF *be any partial $SAF = \langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ defined by an $AT = ((\mathcal{L}, \mathcal{R}, n), \mathcal{K}_n \cup \mathcal{K}_p)$ such that*

1. $\mathcal{L} = \texttt{Prop}(\mathcal{A})$;
2. $\mathcal{K}_n = \emptyset$;
3. $\mathcal{K}_p = \mathcal{A} \cap \mathcal{L}$;
4. $\mathcal{R}_\texttt{s}(\mathcal{A}) = \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi \mid$ *there exists an* $A = [A_1], \ldots, [A_n] \rightarrow \psi \in \mathcal{A}\}$;

5. $\mathcal{R}_{\mathtt{d}}(\mathcal{A}) = \{\mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \Rightarrow \psi \mid$ *there exists an* $A = [A_1], \ldots, [A_n] \Rightarrow \psi \in \mathcal{A}\}$*;*
6. n *is any partial function from* $\mathcal{R}_d$ *to* $\mathcal{L}$*;*
7. $\preceq$ *is any ordering on* $\mathcal{A}$*.*

This definition is in fact a fixpoint construction and it may have multiple fixpoints, as will be shown below. The idea is that it formally reconstructs the set $\mathcal{A}$ of arguments stated in the debate and adds no further arguments but that the argument preferences are provided by an evaluator of the debate. As for the n function it is simply assumed that it can be sensibly identified from the debate, without going into details. The following proposition states that the construction of Definition 8 is well-defined.

Proposition 1. *For any debate-generated* $SAF = \langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ *defined by an* $AT = ((\mathcal{L}, \mathcal{R}, n), \mathcal{K}_n \cup \mathcal{K}_p)$ *it holds that all elements of* $\mathcal{A}$ *are constructible on the basis of* $\mathcal{K}_n \cup \mathcal{K}_p$ *in* $(\mathcal{L}, \mathcal{R}, n)$*.*

Proof. The proof is by induction on the construction of arguments. If $A \in \mathcal{L}$ then $A \in \mathcal{K}_p$ by clause (3) of Definition 8, so A is constructible. Otherwise A is of the form $A = [A_1], \ldots, [A_n] \rightarrow/\Rightarrow \psi \in \mathcal{A}\}$. By the induction hypothesis, $A_1, \ldots, A_n$ are constructible. Moreover, $\mathtt{Toprule}(A)$ is in $\mathcal{R}_s$ or $\mathcal{R}_d$ by clause (4) or (5) of Definition 8. So A is constructible.

Corollary 1. *For any debate-generated* SAF *the set* $\mathcal{A}$ *is closed under the subargument relation.*

The converse of Proposition 1 does not hold in general. Consider the following debate:

A_1: p $\quad$ B_1: s
A_2: $A_1 \Rightarrow q$ $\quad$ B_2: $B_1 \Rightarrow q$
A_3: $A_2 \Rightarrow r$ $\quad$ B_3: $B_2 \Rightarrow t$

Then the following argument C is constructible on the basis of the debate-generated SAF:

A_1: p
A_2: $A_1 \Rightarrow q$
C: $A_2 \Rightarrow t$

However, C is not in $\mathcal{A}$. This also shows that the construction of Definition 1 may have multiple fixpoints, since there exist a debate-generated SAF with $\mathcal{A} = \{A_1, A_2, A_3, B_1, B_2\}$ and one with $\mathcal{A} = \{A_1, A_2, A_3, B_1, B_2, C\}$.

Now the problem motivating this paper arises if for some arguments $A_i \in \mathcal{A}_i$ and $A_j \in \mathcal{A}_j (i \neq j)$ we have that A_i's conclusion is an ordinary premise of A_j, where neither argument is atomic (so not an element of $\mathcal{L}$). For example, consider a debate variant of a well-known example from the literature on nonmonotonic logic, with a propositional language. John says "Nixon was a pacifist (p) since he was a Quaker (q) and Quakers are usually pacifists ($q \Rightarrow p$)", while Bob says "Nixon was not a pacifist ($\neg p$) since he was a republican (r) and republicans are usually not pacifists ($r \Rightarrow \neg q$)" (thus generalisations are modelled as defeasible inference rules). Formally:

A_1: q $\quad$ B_1: r
A_2: $A_1 \Rightarrow p$ $\quad$ B_2: $B_2 \Rightarrow \neg p$

Suppose now that Mary supports John's argument by saying "Nixon regularly attended service in a Quaker church (c), people who are regularly seen in a Quaker church usually are a Quaker ($c \Rightarrow q$), so Nixon was a Quaker". Formally:

A_3: c
A_4: $A_3 \Rightarrow q$

It is this kind of situation that is of special interest in this paper. On the one hand, we want to capture that in some sense argument A_4 supports argument A_2. For example, the evaluator might not be prepared to accept q if it is not supported by some argument. On the other hand, we do not want to simply replace these two arguments with the following combined argument, while deleting q from $\mathcal{K}$ of the debate-generated SAF:

A_3: c
A_4: $A_2 \Rightarrow q$
A_5: $A_4 \Rightarrow p$

The reason why we do not want to do this is that there can also be situations in which the evaluator does not accept argument A_5 for q but still accepts q as a premise: so argument A_5 should be part of the debate on its own, at least initially.

This paper's solution will be that both the two individual arguments A_2 and A_4 and their combination A_5 will initially be part of the debate and that the evaluator should for each premise of A_2 decide whether to accept it without further argument. If so, then A_5 is irrelevant for the issue p whether Nixon was a pacifist, otherwise, A_2 will be removed from the debate and A_5 becomes relevant. It is this approach that will be formalised in Sections 5 and 6. But first the alternative solution presented by [3] will be discussed.

4 The BAF approach with premise support

In *ASPIC*$^+$ the only support relation between arguments is the subargument relation as defined in Definition 3, where each argument contains all its subarguments as part of itself. Let us now examine Cayrol & Lagasquie-Schiex's claim in [3] that to model debates, instead support relations between separate arguments are needed. To this end, consider a version of *ASPIC*$^+$ that generates SuppAFs, leaving the definitions of conflict-freeness and admissibility as they are. Two ways of defining the support relation suggest themselves: the original subargument relation from *ASPIC*$^+$ and the definition that A supports B if the conclusion of A is a premise of B (henceforth called *premise support*). *ASPIC*$^+$-SuppAFs with subargument support were in [21], building on [20], shown to be equivalent to *ASPIC*$^+$ when generating AFs as defined above in Definition 7 (this result was reproved by [4]). However, the same does not hold for *ASPIC*$^+$-SuppAFs with premise support. The crucial difference between premise support and ASPIC's subargument relation is that while each argument contains all subarguments as part of itself, premise-support relations can hold between different arguments of which neither contains the other.

To illustrate the difference, consider again the above modelling the Nixon example. Argument A_5 for the conclusion that Nixon was a pacifist contains the arguments A_3 that Nixon was regularly seen in a Quaker church and A_4 that therefore he was a pacifist as part of itself. This modelling is by Cayrol & Lagasquie-Schiex considered less natural. They want that John's, Bob's and Mary's arguments are three individual arguments, which all stand on their own but can be related by support relations. For instance, John's argument has a premise q and conclusion p and is supported by Mary's argument which has premise c and conclusion q: John's argument does not contain Mary's argument as part of itself. Likewise, Bob's argument stands on its own, with premise r and conclusion $\neg p$.

Let us see how Definition 8 can be applied to respect this view. There exists a debate-generated SAF which only contains the arguments $A_1, A_2, A_3, A_4, B_1, B_2$ and not A_5. This (partial) SAF is defined as follows.

- $\mathcal{L} = \{c, p, q, r, \neg p\}$;
- $\mathcal{K}_p = \{c, q, r\}$;
- $\mathcal{R}_{\mathtt{s}}(\mathcal{A}) = \emptyset$;
- $\mathcal{R}_{\mathtt{d}}(\mathcal{A}) = \{c \Rightarrow q; q \Rightarrow p; r \Rightarrow \neg p\}$.
- $n = \emptyset$;
- $\preceq$ is any;
- $\mathcal{A} = \{A_1, A_2, A_3, A_4, B_1, B_2\}$.

With premise support we have that the support relations are that A_1 supports A_2, B_1 supports B_2, A_3 supports A_4 and A_4 supports A_2.

Note that Definition 8 also allows a debate-generated SAF with q deleted from $\mathcal{K}_p$ and A_5 added to $\mathcal{A}$ (recall that Definition 8 is a fixpoint construction that may have multiple fixpoints). However, this is not a problem, since the point is that an evaluator can, following [3], decide that the first debate-generated SAF is the one corresponding to the debate s/he is analysing.

Is this then the way *ASPIC*$^+$ can be used for evaluating debates, by letting it generate SuppAFs as just sketched with the notion of premise support? The first issue here is whether defeating a supporter of an argument should have an impact on the supported argument. So far this is with premise support not guaranteed. Suppose in the Nixon example that argument A_3 is defeated by an argument C. Then C does not indirectly defeat A_1 or A_2, since A_3 is not a subargument of these arguments (recall that A_5, of which A_3 is a subargument, is not in the SuppAF with premise support). So in SuppAFs as defined thus far, defeating a supporter of an argument does not have any logical effect on the status of the supported argument.

For contexts in which arguments are generated from a given knowledge base this is clearly undesirable. However, at first sight, it might be argued that for debate contexts this is otherwise. Suppose that Bob in our example debate succeeds in defeating Mary's argument by arguing that Nixon only attended service in a Quaker church to please his wife, who was a Quaker. It could reasonably be argued that this does not knock down John's argument, since why should John be blamed for Mary's flawed attempt to support his argument? On the other hand, it is still unsatisfactory that there is no logical relation at all between defeating a supporter and the status of a supported argument. If support means anything at all, then surely defeating a supporter should have some logical effect on the status of an argument supported by it. The problem then is how this view can be reconciled with the view that in debate contexts this logical effect cannot simply be that the supported argument is always defeated.

At first sight, a possible solution would be to close the defeat relation in a SuppAF under the constraint (adopted from [8, 16]) that if A supports B and C defeats A, then C also defeats B. This constraint (in the literature on BAFs called 'secondary attack') is both necessary and sufficient to prove that *ASPIC*$^+$-SuppAFs with subargument support are equivalent to *ASPIC*$^+$-$SAFs$ [21]. Can *ASPIC*$^+$-SuppAFs with premise support and with secondary attacks be used for evaluating debates? This is still not true, for two reasons. The first reason is that then argument B is (indirectly) defeated by C regardless whether the evaluator wants to accept its premise q without further support. The second reason was given in [21], namely, that such SuppAFs cannot distinguish between the following two situations:

> **Situation 1**: A has premises p and q, B has conclusion p, C has conclusion q, D undercuts C.
> **Situation 2**: A has premise p and both B and C have conclusion p, D undercuts C.

Both situations induce the same SuppAF with premise support and secondary attack, in which both B and C support A and D defeats both A and C. However, this is counterintuitive, since in the second situation A should not be defeated, since its premise p is still provided by an undefeated argument, namely, B. With support as subargument support instead the intuitive outcome is obtained, since then there are in situation 2 two 'versions' of A, one with B as subargument for q and the other with C as subargument for q; and only the second version of A is defeated by D. It can be concluded that the notion of secondary attack is not suitable for premise support.

More generally, this analysis shows that support relations between arguments in debates after all have to be modelled by combining the two arguments into a single complex one, that is, as *ASPIC*$^+$-style subargument relations. But how can this be done in a way that still respects that the combined arguments may have been stated by different debate participants? This paper's solution to this problem will be based on the idea that debates are always evaluated relative to a given or assumed 'basis for discussion' (an idea also underlying the Carneades framework of [10]). This basis can be objective or authoritative, such as when a judge or jury in a legal proceeding has to evaluate a dispute between opposing parties, but it can also be purely subjective, such as when an ordinary citizen evaluates a political debate in light of his or her subjective opinions. In the Nixon example, if the evaluator accepts the premise of John's argument that Nixon was a Quaker, then defeating Mary's supporting argument has no effect on the status of John's argument. But if the evaluator does not accept this premise but only accepts Mary's premise that Nixon regularly attended service in a Quaker church, then defeating Mary's argument also defeats John's argument. To make this work, both the original arguments and their combined version should be considered in the evaluation process.

5 Combining arguments

Recall that in debate evaluation we are not interested in the sources of the arguments but only in their logical and dialectical relations. We therefore face the problem of combining arguments that may come from different sources. Since, as just shown, in the present formal context the notion of premise support cannot be used for this purpose but the *ASPIC*$^+$ subargument relation has to be used. The problem now is how to do this, that is, how to extend debate-generated $SAFs$ as defined in Definition 8 with further arguments and attack relations.

The idea is twofold. First, if the conclusion of an argument A provides a premise of an argument B, then the arguments are combined into a new argument by replacing the premise in B by A as a subargument of B. This is in fact the notion of 'backwards extending' an argument from [18] and the notion of 'weakening' an argument from [6]. But this is not all. It may happen that an argument A has the same conclusion as a non-premise proper subargument B' of B. For example, suppose argument A_5 in our Nixon example was stated by a debate participant and now another participant states a new argument with conclusion q, for instance, that an inhabitant of Nixon's home town says that Nixon was a Quaker (h).

A_6: h
A_7: $A_6 \Rightarrow q$

Then it seems reasonable to create a new version of A_5 with its subargument A_4 replaced by A_7. It turns out that both cases (see Figure 1) can be combined into a single inductive definition.

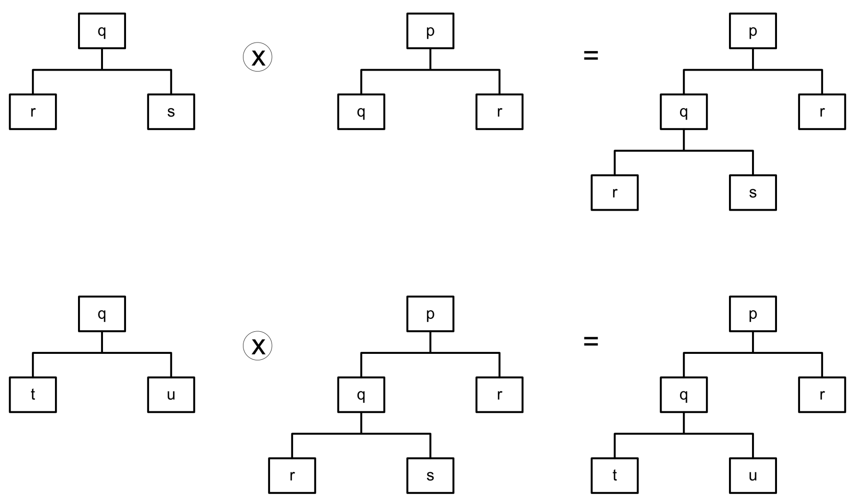

Fig. 1. Two ways to modify an argument

Definition 9. *For any argument B and wff φ, $B^\varphi = \varphi$ iff $\texttt{Conc}(B) = \varphi$, otherwise $B^\varphi = B$.*

For any pair of arguments $A \notin \mathcal{L}$ and B such that $A \notin \texttt{Sub}(B)$ the modification *$A \otimes B$ of B by A is inductively defined as follows:*

1. *if $\texttt{Conc}(A) \neq \texttt{Conc}(B')$ for all $B' \in \texttt{Sub}(B)$ then $A \otimes B = B$, else:*

2. *if $B \in \mathcal{L}$ then $A \otimes B = A$;*
3. *if $B \notin \mathcal{L}$ then we have that B is of the form $B_1, \ldots, B_n \to / \Rightarrow \psi$: then $A \otimes B = A \otimes B_1^{\varphi}, \ldots, A \otimes B_n^{\varphi} \to / \Rightarrow \psi$, where $\varphi = \texttt{Conc}(A)$.*

In case (1) A does not add anything to B so B remains unchanged. In case (2) B is an item from the knowledge base which is equal to A's has conclusion; then A replaces B. Case (3) is the inductive clause. For any $B_i(1 \leq i \leq n)$ such that A's conclusion equals the conclusion of B_i, B_i is replaced with A in B by first replacing B_i with A's conclusion and then applying case (2) to the resulting premise argument. For all remaining $B_j(j \neq i)$ case (3) is inductively applied. In sum, case (2) replaces premises of B with an argument for the premise (the first case in Figure 1) while case (3) replaces subarguments of B with alternative subarguments for the same conclusion (the second case in Figure 1) and takes care of the recursion.

Definition 10. *The* closure *of a set $\mathcal{A}$ of arguments is the smallest set $\mathcal{A}^{\otimes}$ such that:*

1. *If $A \in \mathcal{A}$ then $A \in \mathcal{A}^{\otimes}$;*
2. *If $A \in \mathcal{A}^{\otimes}$ and $B \in \mathcal{A}^{\otimes}$ then $A \otimes B \in A^{\otimes}$.*

A set $\mathcal{A}$ of arguments is closed *if $\mathcal{A}^{\otimes} = \mathcal{A}$.*

A debate-generated SAF with a closed set of arguments is always a non-partial SAF.

Proposition 2. *For any debate-generated $SAF = \langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ defined by an AT where $\mathcal{A}$ is closed, it holds that $\mathcal{A}$ is the set of all arguments on the basis of AT.*

Proof. The only-if part (all arguments in $\mathcal{A}$ are constructible on the basis AT) is Proposition 1, while the if-part (all arguments constructible on the basis of AT are in $\mathcal{A}$) is proven as follows. If A is constructible since $A \in \mathcal{K}_p$, then $A \in \mathcal{A} \cap \mathcal{L}$ by clause (3) of Definition 8, so $A \in \mathcal{A}$. Then if A is constructible since $A_1, \ldots, A_n$ are constructible and $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \to/\Rightarrow \psi \in \mathcal{R}_s/\mathcal{R}_d$ then by the induction hypothesis $A_1, \ldots, A_n$ are in $\mathcal{A}$. By clause (4) or (5) of Definition 8 there exists an argument $A' = A'_1, \ldots, A'_n \to/\Rightarrow \psi$ in $\mathcal{A}$. Note that $A \otimes A' = A_1, \ldots, A_n \to/\Rightarrow \psi$. Then $A \in \mathcal{A}$ by clause (3) of Definition 9 and clause (2) of Definition 10.

Looking back at the counterexample to the if-part of Proposition 1, it can be seen that it is now excluded since $C = A_2 \otimes B_3$.

Some but in general not all attack and defeat relations are preserved under $\otimes$-closure.

Proposition 3. *For any triple of arguments A, B and C it holds that*

1. *if C attacks A then C attacks $A \otimes B$;*
2. *if $A \otimes B$ only replaces a premise of B with A, then if C rebuts or undercuts B then C attacks $A \otimes B$;*
3. *if C undermines B and A has a defeasible top rule, then C rebuts $A \otimes B$;*
4. *if C rebuts or undercuts B on its subargument B' and B' is also a subargument of $A \otimes B$, then C also rebuts or undercuts $A \otimes B$ on B';*
5. *if C defeats A then C defeats $A \otimes B$.*

Proof. (1) If C undermines A on φ then C undermines $A \otimes B$ on φ since by construction all premises of A are also premises of $A \otimes B$. If C rebuts or undercuts A on its subargument A' then C also rebuts or undercuts $A \otimes B$ on A' since by construction A' is also a subargument of $A \otimes B$.

(2) If C rebuts or undercuts B on its subargument B' then C also rebuts or undercuts $A \otimes B$ on B' since by construction A' is also a subargument of $A \otimes B$.

(3) Suppose C undermines B on φ. Then A has a defeasible top rule with consequent φ so C rebuts $A \otimes B$ on A.

(4) obvious.

(5) If C undermines A on φ then C undermines $A \otimes B$ on φ since by construction all premises of A are also premises of $A \otimes B$. Then C defeats B since $C \not\prec \varphi$. If C undercuts A then C also undercuts $A \otimes B$ by (1), so C defeats $A \otimes B$. Finally, if C rebuts A on A' then C also rebuts $A \otimes B$ on A' by (1). Then C defeats $A \otimes B$ since $C \not\prec A'$.

Properties (2-4) do not in general hold for defeat, since in general an argument may become 'weaker'[4] according to $\preceq$ if one of its premises or non-premise subarguments is replaced by another argument. For example, with [13]'s weakest-link ordering the following may happen (with the rule names attached to $\Rightarrow$ for clarity).

A_1: r B_1: p D_1: s
A_2: $A_1 \Rightarrow_{r1} p$ B_2: $B_1 \Rightarrow_{r2} q$ D_2: $D_1 \Rightarrow_{r3} \neg q$
C: $A_2 \Rightarrow_{r2} q$

If we have that $r_1 < r_3 < r_2$ then B defeats D but $C = A \otimes B$ does not defeat D. Even with [13]'s last-link ordering properties (2-4) do not in general. For example, if $A \otimes B$ replaces a non-premise subargument B' of B with A and C rebuts B on B', then it may be that $C \not\prec B'$ but $C \prec B$. Likewise if B' is a premise of B.

6 Applying the definitions to debate evaluation

The approach proposed in the previous section solves the problems with the approach discussed in Section 4 to generate SuppAFs. When an argument in a debate has multiple supports on the same premise, it will be multiplied as desired. The following example debate illustrates this. Let $\mathcal{A} = \{A_1, A_2, B_1, B_2, C_1, C_2\}$ where:

A_1: p B_1: r C_1: s
A_2: $A_1 \Rightarrow q$ B_2: $B_1 \Rightarrow p$ C_2: $C_1 \Rightarrow p$

Then $\mathcal{A}^{\otimes}$ adds the following arguments to $\mathcal{A}$:

D_1: $B_1 \Rightarrow p$ E_1: $C_1 \Rightarrow p$
D_2: $D_1 \Rightarrow q$ E_2: $E_1 \Rightarrow q$

Here:

$D_1 = B_2 \otimes A_1$
$D_2 = D_1 \otimes A_2 = (B_2 \otimes A_1) \otimes A_2$
$E_1 = C_2 \otimes A_1$
$E_2 = E_1 \otimes A_2 = (C_2 \otimes A_1) \otimes A_2$

[4] In this context an argument B is said to be 'weaker' than an argument A (and A 'stronger' than B) if there exists an argument C such that $B \prec C$ but not $A \prec C$.

So $\mathcal{A}^{\otimes}$ contains three different arguments for q: the original argument A_2 from $\mathcal{A}$ and the new arguments D_2 and E_2.

Now an evaluator who accepts statement p without further argument, can modify the AT of the SAF by moving p from $\mathcal{K}_p$ to K_n. If, furthermore, the argument ordering is such that an argument can never be 'strengthened' by replacing a necessary premise with a non-premise subargument, then the two arguments for p become irrelevant to the issue q and only argument A_2 counts for this issue. By contrast, an evaluator who does not accept p without further argument can delete p from $\mathcal{K}$ after which A_2 is not constructible any more and the arguments D_2 and E_2 become relevant to the issue q. (An alternative approach is to add 'issue premises' to *ASPIC*$^+$ as in [19] and to move p to the issue premises). This in fact embodies a dynamic view on debate evaluation, where evaluators can not only provide preferences for given arguments, but can also modify or discard arguments and perhaps add arguments of their own. This is arguably a realistic view on debate evaluation.

Let us briefly discuss some alternative ways to define the closure of a set of arguments resulting from a debate, which further illustrate that debate evaluation need not constrain itself to given arguments. Consider first the just-given example. The original set $\mathcal{A}$ contains two arguments for p: one with r and the other with s as a defeasible reason for p. An evaluator of the debate might then wish to aggregate these reasons in a new argument for p based on a rule $q, r \Rightarrow p$.

Second, a debate might be evaluated against the background of some assumed set of inference rules which were not necessarily used in the debate to construct arguments; that is, they need not be part of the AT of a debate-generated SAF. For example, a set of strict rules generated by a monotonic logic for $\mathcal{L}$ might be assumed (such as classical logic) or a set of defeasible argument schemes might be assumed. In these cases the closure of a set $\mathcal{A}$ of arguments might be defined to contain 'implied' arguments. For example, if $\mathcal{A}$ contains $A = p \Rightarrow q$ and $B = r \Rightarrow s$ and the evaluator assumes a rule $p, r \Rightarrow \neg q$, then $\mathcal{A}^{\otimes}$ might be defined to also contain $C = p, r \Rightarrow \neg q$, which attacks A even though A was not actually attacked in the debate.

7 Conclusion

In this paper a formal model was proposed of support relations between arguments put forward in debates in the context of the theory of abstract argumentation frameworks [5]. This context was chosen in order to evaluate claims from the literature on bipolar argumentation frameworks that just having attack relations between arguments would be insufficient for modelling debates. We learned that there are problems with the idea that arguments that support each other can be evaluated as independent entities and that it is better to combine supporting and supported arguments in a single compound argument by using *ASPIC*$^+$'s subargument relation. A method was defined to formally reconstruct the set of arguments put forward in a debate in a way that respects these observations. Then several ways were sketched in which debates can be evaluated in terms of the formal reconstruction. An important insight was that to correctly model support relations between arguments put forward in debates, an explicit formal account was needed of the structure of arguments and the nature of attack and defeat relations between arguments. This casts doubt on the applicability of purely abstract models of debate evaluation.

This paper thus also provides an answer to Betz's criticism in [1] that the theory of abstract argumentation frameworks would not be suitable for modelling debates. This criticism might be justified if these frameworks are applied on their own, but in this paper we have seen that if they are combined with accounts of the structure of argumentation and of how debates with structured arguments can be evaluated, then the theory of abstract argumentation frameworks is still a useful component of an adequate model of debate evaluation.

As for related research, this paper's approach is most closely related to the Carneades system of [10, 11], which was originally proposed as a formalism for evaluating debates. The present idea to let the evaluation of arguments partly depend on whether the evaluator accepts its premises without further support was taken from the 2007 version of Carneades. A main difference in approach between $ASPIC^+$ (and other instantiations of the theory of abstract argumentation frameworks) on the one hand and Carneades (and also e.g. [2]'s Abstract Dialectical Frameworks, ADFs) on the other is that while in $ASPIC^+$ the main focus is on evaluating *arguments*, in Carneades it is on evaluating *statements*. Because of this difference, the issues discussed in this paper do not arise in the same way in Carneades or ADFs. Future research should shed further light on the relative merits of both approaches as regards these issues.

Acnowledgement

I thank Tom Gordon for his comments on an earlier version of this paper.

References

1. G. Betz. Evaluating dialectical structures. *Journal of Philosophical Logic*, 38:283–312, 2009.
2. G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference*, pages 102–111. AAAI Press, 2010.
3. C. Cayrol and M.-C. Lagasquie-Schiex. Bipolar abstract argumentation systems. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 65–84. Springer, Berlin, 2009.
4. A. Cohen, S. Parsons, E. Sklar, and P. McBurney. A characterization of types of support between structured arguments and their relationship with support in abstract argumentation. *International Journal of Approximate Reasoning*, 94:76–104, 2018.
5. P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and *n*–person games. *Artificial Intelligence*, 77:321–357, 1995.
6. P.M. Dung. An axiomatic analysis of structured argumentation with priorities. *Artificial Intelligence*, 231:107–150, 2016.
7. P.M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171:642–674, 2007.
8. P.M. Dung and P.M. Thang. Closure and consistency in logic-associated argumentation. *Journal of Artificial Intelligence Research*, 49:79–109, 2014.

9. A.J. Garcia and G.R. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4:95–138, 2004.
10. T.F. Gordon, H. Prakken, and D.N. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171:875–896, 2007.
11. T.F. Gordon and D.N. Walton. Formalizing balancing arguments. In P. Baroni, T.F. Gordon, T. Scheffler, and M. Stede, editors, *Computational Models of Argument. Proceedings of COMMA 2016*, pages 327–338. IOS Press, Amsterdam etc, 2016.
12. N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical-logic arguments: postulates and properties. *Artificial Intelligence*, 175:1479–1497, 2011.
13. S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.
14. S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation*, 5:31–62, 2014.
15. S. Modgil and H. Prakken. Abstract rule-based argumentation. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*, volume 1, pages 73–141. College Publications, London, 2018.
16. F. Nouioua and V. Risch. Argumentation frameworks with necessities. In *Proceedings of the 4th International Conference on Scalable Uncertainty Management (SUM'11)*, number 6929 in Springer Lecture Notes in AI, pages 163–176, Berlin, 2011. Springer Verlag.
17. J.L. Pollock. Justification and defeat. *Artificial Intelligence*, 67:377–408, 1994.
18. H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.
19. H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124, 2010.
20. H. Prakken. Relating ways to instantiate abstract argumentation frameworks. In K.D. Atkinson, H. Prakken, and A.Z. Wyner, editors, *From Knowledge Representation to Argumentation in AI, Law and Policy Making. A Festschrift in Honour of Trevor Bench-Capon on the Occasion of his 60th Birthday*, pages 167–189. College Publications, London, 2013.
21. H. Prakken. On support relations in abstract argumentation as abstractions of inferential relations. In *Proceedings of the 21st European Conference on Artificial Intelligence*, pages 735–740, 2014.
22. G.R. Simari and R.P. Loui. A mathematical treatment of defeasible argumentation and its implementation. *Artificial Intelligence*, 53:125–157, 1992.