Formalising an Aspect of Argument Strength: Degrees of Attackability

Henry PRAKKEN

Department of Information and Computing Sciences, Utrecht University and Faculty of Law, University of Groningen, The Netherlands and European University Institute, Florence, Italy

Abstract. This paper formally studies a notion of dialectical argument strength in terms of the number of ways in which an argument can be successfully attacked in expansions of an abstract argumentation framework. The proposed model is abstract but its design is motivated by the wish to avoid overly limiting assumptions that may not hold in particular dialogue contexts or in particular structured accounts of argumentation. It is shown that most principles for gradual argument acceptability proposed in the literature fail to hold for the proposed notion of dialectical strength, which clarifies their rational foundations and highlights the importance of distinguishing between logical, dialectical and rhetorical argument strength.

Keywords. Dialectical argument strength, Structure of arguments, Nature of attack

1. Introduction

A recent trend in the formal study of argumentation is the development of gradual notions of argument acceptability, as alternatives to extension-based notions defined on top of the theory of abstract [9] or bipolar [7] argumentation frameworks. In [14] we argued that such work should make explicit which kind of argument strength or acceptability is modelled, since different kinds of strength may have different properties. In particular, we distinguished between logical, rhetorical and dialectical argument strength.

Logical argument strength in turn divides into two aspects. *Inferential argument strength* is about how well an argument's premises support its conclusion considering only the argument itself. For example, deductive arguments are stronger than defeasible arguments. *Contextual argument strength* is about how well the conclusion of an argument is supported in the context of all given arguments. Formal frameworks like Dung's theory of abstract argumentation frameworks, assumption-based argumentation, $ASPIC^+$ and defeasible logic programming formalise this kind of argument strength [10].

Rhetorical argument strength looks at how capable an argument is to persuade other participants in a discussion or an audience. Persuasiveness essentially is a psychological notion; although principles of persuasion may be formalised, their validation as principles of successful persuasion is ultimately psychological.

Finally, *dialectical argument strength* looks at how challengeable an argument is in the context of a critical discussion. In [15, pp. 657] this is formulated as

> (...) the (un)availability of participant moves that constrain further interlocutor moves. Minimally, argument strength thus is a function of the (un)availability of non-losing future participant moves. In this sense, the strongest proponent-argument leaves no further opponent-move except concession (i.e., retraction of either a standpoint or of critical doubt), and the weakest proponent argument constrains no opponent-move, given the "move-space".

Thus conceived, an important aspect of dialectical strength is the degree of attackability of an argument, that is, how many attacks are allowed in a given state that decrease the argument's contextual status. This reflects an intuition that many decision makers are aware of, namely, to justify one's decisions as sparsely as possible, in order to minimise the chance of successful appeal. It is this notion of dialectical strength that is the focus of the present paper.

We first propose a refined version of the notion of a normal expansion [3] of an abstract argumentation framework, designed so as to avoid overly limiting assumptions about the nature of arguments and their relations and the dialogical context. We then formalise dialectical argument strength in terms of teh number of ways to expand an argumentation framework such that the argument's contextual status decreases. We define this notion in two equivalent ways (ranking-based and weighted) and we investigate some of its formal properties. Among other things, we show that most principles for gradual argument acceptability proposed in the literature fail to hold for our notion of dialectical strength, which says something about the rational foundations of these principles.

2. Formal Preliminaries

An *abstract argumentation framework* (AF) [9] is a pair $(\mathcal{A}_{AF}, \mathcal{C}_{AF})$, where $\mathcal{A}_{AF}$ is a set of arguments and $\mathcal{C}_{AF} \subseteq \mathcal{A}_{AF} \times \mathcal{A}_{AF}$ is a relation of attack. We write $A \in AF$ as shorthand for $A \in \mathcal{A}_{AF}$ and we will omit the subscripts if there is no danger for confusion. We will sometimes in text present an AF as $A \leftarrow B \leftrightarrow C$, to denote that $\mathcal{A} = \{A, B, C\}$ and $\mathcal{C} = \{(B, A), (B, C), (C, B)\}$. Argument A is an *attacker* of argument B if $(A, B) \in \mathcal{C}$, and A is a *direct defender* of B if for some attacker C of B it holds that $(A, C) \in \mathcal{C}$. An *attack branch*, respectively, *defense branch* of an argument A_1 is a finite sequence $A_1, \ldots, A_n$ such that n is even, respectively, odd, and in both cases A_n has no attackers and for each $i < n$ it holds that A_{i+1} attacks A_i. Argument B_i is a *defender* of argument A_1 iff B_i is in an attack or defense branch of A_1 and $i > 1$ and i is odd.

The semantics of AFs [9,2] identifies sets of arguments (called *extensions*) which are internally conflict-free (no member attacks a member) and defend themselves against all attackers. In this paper we use the labelling way to define semantics for AFs. A *labelling* of a set $\mathcal{A}$ of a set of arguments in an $AF = (\mathcal{A}, \mathcal{C})$ is any triple of non-overlapping subsets (*in*,*out*,*und*) of $\mathcal{A}$ that satisfies the following constraints:

1. an argument is *in* iff all arguments attacking it are *out*;
2. an argument is *out* iff it is attacked by an argument that is *in*;
3. an argument is *und* (for 'undecided') iff it is neither *in* nor *out*.

In this paper we focus on grounded semantics, leaving generalisation to other semantics for future research. The grounded labelling of an AF minimises the set of arguments that are labelled *in* and is always unique. A set $S \subset \mathcal{A}$ is called the *grounded extension* of AF iff S is the set of all arguments labelled *in* in the grounded labelling.

3. Dialectical Argument Strength: ranking-Based Semantics

In this section we define a ranking-based semantics of dialectical strength of arguments in the form of a preorder on the set of arguments. Dialectical argument strength has both static and dynamic aspects. A static aspect is whether an argument has been successfully defended in a terminated dialogue, which is a matter of applying a notion of contextual strength at termination. Dynamic aspects concern how challengeable an argument is in a given non-final state of the dialogue. Taking the formulation of [15] quoted above in the introduction literally, it should be modelled by considering all possible ways to terminate the dialogue but in general this is infeasible sicne it will often be impossible to foresee which information is available to construct arguments, how they will be evaluated, and which procedural decisions (such as on admissibility of evidence) will be taken.

For these reasons, we propose the following approach. Imagine a dialogue participant who can extend a given AF and who wants to make a given argument F (the focus argument) dialectically as strong as possible. The participant will consider all procedurally allowed expansions AF' of AF and determine in which of these expansions F is the strongest. So in general we have to compare arguments that are in *different* AFs. Moreover, our notion of strength will not boil down to applying a notion of contextual strength to all these expansions, since we also want to determine how vulnerable F is to attack in all these expansions. To this end we will define a notion of 'attack points' of an argument, which are minimal sets of arguments that, if attacked in an allowed expansion, make the contextual status of the focus argument decrease.

To model these ideas, we let dialectical strength be determined by a combination of the 'current' contextual strength of an argument and its number of attack points as follows. To start with, we assume a ranking of contextual argument statuses, which in the present paper will be that being labelled *in* is better than being labelled *undecided*, which is better than being labelled *out*. In notation: *in* $>_c$ *und* $>_c$ *out*. (In future research this could be extended to alternative semantics, even to gradual ones, but in this paper we prefer to keep things simple to focus on the essence.) Then, given the set of allowed expansions $\{AF', AF'', \ldots\}$ of a given AF, we say that if argument $A_{AF'}$ is contextually better than argument $B_{AF''}$ then it is also dialectically better than argument $B_{AF''}$, while if $A_{AF'}$ and $B_{AF''}$ are contextually equally strong, then $A_{AF'}$ is better than $B_{AF''}$ if $A_{AF'}$ has fewer attack points than $B_{AF''}$. So this notion of dialectical strength presupposes and is a refinement of the notion of contextual strength. The primacy of contextual strength is justified by our intended application scenario, where a proponent of a focus argument F wants to move to a state where F is contextually as strong as possible. Moreover, if contextual strength has primacy, then for terminated disputes dialectical strength reduces as desired to how well an argument is defended at termination.

Consider an example $AF = A \leftarrow B$ and let A be the focus argument. Assume the proponent of A can expand AF with either C, resulting in $AF' = A \leftarrow B \leftarrow C$, or with D, resulting in $AF'' = A \leftarrow B \leftarrow D$. In both expansions A is *in* so contextually of the same strength. However, assume that C is attackable while D is unattackable. Then A has two attack points in AF', namely, $\{A\}$ and $\{C\}$, while A has only one attack point in AF'', namely, $\{A\}$. So A is dialectically stronger in AF'' than in AF', so the dialectically better choice for the proponent is to expand AF to AF'' by moving D.

An attack point must be defined as a *set* of arguments. Consider Figure 1. Attacking just C or just D is not enough to lower the status of A, so one attack point must in this

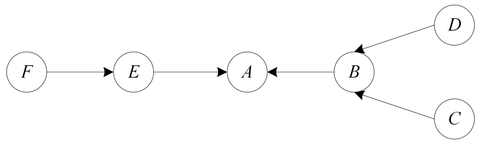

Figure 1. Multiple attack points

case be defined as $\{C, D\}$. Note, furthermore, that attacking F also lowers the status of A, so $\{F\}$ is also an attack point of A, so an argument can have multiple attack points.

To define attack points, we now first define the notion of an allowed expansion of an AF, which is a refinement of [3]'s notion of a normal expansion. The first refinement is to make expansions relative to a given background universal argumentation framework $UAF = (\mathcal{A}^u, \mathcal{C}^u)$. An important reason for doing so is to avoid implicit assumptions at the abstract level that are not always satisfied by instantiations, such as that all arguments are attackable or that all attacks are independent from each other.

Definition 1 Given a universal argumentation framework $UAF = (\mathcal{A}^u, \mathcal{C}^u)$, an *argumentation framework in* UAF is any $AF = (\mathcal{A}, \mathcal{C})$ such that $\mathcal{A} \subseteq \mathcal{A}^u$ and $\mathcal{C} \subseteq \mathcal{C}^u_{|\mathcal{A} \times \mathcal{A}}$.

That $\mathcal{C}$ is not required to equal $\mathcal{C}^u_{|\mathcal{A} \times \mathcal{A}}$ is to allow for instantiations like *ASPIC*$^+$ that use preferences to resolve attacks into defeat relations and let $\mathcal{C}$ stand for defeat.

We must also distinguish between allowed and not allowed expansions. One reason is that the dialogical protocol may impose constraints, such as admissibility of premises or of types of arguments (for example, in some systems of criminal law analogical applications of criminal provisions are not allowed). The problem context may also impose restrictions. For example, investigation procedures in which information gathering is interchanged with argument construction may have a constraint that all and only relevant arguments constructible from the gathered information are included. Finally, underlying structured accounts of argumentation may impose such constraints, for example, a closure constraint on the set $\mathcal{A}'$ of new arguments in that other arguments that can be constructed with information introduced by arguments in $\mathcal{A}'$ must also be in $\mathcal{A}'$.

We now define (allowed) expansions relative to a given UAF as follows.

Definition 2 **[Expansions given a universal argumentation framework]** Let AF = $(\mathcal{A}, \mathcal{C})$ and AF' be two abstract argumentation frameworks in UAF. Then AF' is an *expansion* of AF given UAF if AF' = $(\mathcal{A} \cup \mathcal{A}', \mathcal{C} \cup \mathcal{C}')$ for some nonempty $\mathcal{A}'$ disjoint from $\mathcal{A}$, such that for all A, B: if $(A, B) \in \mathcal{C}'$ then $A \in \mathcal{A}'$ or $B \in \mathcal{A}'$.

Let UAF^e be the set of all expansions of some AF given UAF. Then $aUAF^e \subseteq UAF^e$ is the set of *allowed expansions* given UAF.

A further refinement is needed. Imagine two attackable but unattacked arguments A and B such that for both of them expansions exist that lower their status. Then they both have one attack point, namely, $\{A\}$, respectively, $\{B\}$. However, if A has just one attackable premise while B has two, or A uses one defeasible rule while B uses two, then A should still be dialectically stronger than B. Accordingly, we assume that each argument A in a UAF comes with a finite set $t(A)$ of *attack targets* and we assume that each argument B attacking A attacks A on at least one of A's attack targets. Given a set S of arguments, we write S^t for the set of all pairs (A, t) such that $A \in S$ and $t \in t(A)$.

Finally, we need a notion of relevance of a set of defenders to the status of the defended argument. It adapts the dialogical notion of relevance proposed in [13] to AFs.

Definition 3 For any $AF = (\mathcal{A}, \mathcal{C})$ with $A \in \mathcal{A}$, a set $S \subseteq A$ is *relevant to* A in AF iff S is a minimal set such that the contextual status of A is lower in $AF' = (\mathcal{A}', \mathcal{C}')$ than in AF, where $\mathcal{A}' = (\mathcal{A} \cup \{X\})$ for some X not in $\mathcal{A}$ and not attacked by $\mathcal{A}$, and $\mathcal{C}' = \mathcal{C} \cup \{(X, B) \mid B \in S\}$.

So S is relevant to A in AF iff AF can be expanded with an unattacked attacker of all members of S such that A's contextual status is lowered. Note that this notion is not defined relative to a UAF. If S is relevant to A then all arguments in S are defenders of A but it can happen that a defender of A is in no set relevant to A. In Figure 2, C and G are defenders of A but attacking either of them does not lower the status of A; this only happens if either A or D is attacked, so the only sets relevant to A are $\{A\}$ and $\{D\}$.

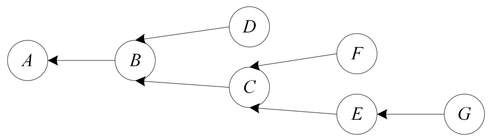

Figure 2. Relevant sets

We are now in the position to define the notion of an attack point of an argument.

Definition 4 [**Attack points**] Given an abstract argumentation framework $AF = (\mathcal{A}, \mathcal{C})$ in UAF, an *attack point* of an argument $A \in \mathcal{A}$ is any minimal set $S \subseteq \mathcal{A}^t$ relevant to A such that an allowed expansion $AF' = (\mathcal{A} \cup \mathcal{A}', \mathcal{C} \cup \mathcal{C}')$ of AF given UAF exists with

1. for all $(B, t) \in S$ there exists an argument $C \in \mathcal{A}'$ such that C attacks B on t;
2. the contextual status of A is lower in AF' than in AF.

The set of attack points of A given AF is denoted by $ap_{AF}(A)$.

It is not required that all arguments in $\mathcal{A}'$ attack some argument in S, since including an attacker of S in $\mathcal{A}'$ might require putting other arguments in $\mathcal{A}'$ as well, such as A's subarguments in systems in which arguments have subarguments. Also, Definition 4 allows for 'side effects' in that the new attackers may also attack arguments outside S or arguments in $\mathcal{A}$ but outside S may attack them. For example, an argument attacking another argument on its premise may also attack all other arguments using that premise.

We can now give our definition of dialectical argument strength, by combining the notion of contextual strength with the number of attack points of arguments. Several definitions are still possible and the ones given by us are not meant to be the final answer but instead to initiate the discussion about what are good definitions. First, we give primacy to the current contextual evaluation in that being contextually stronger implies being dialectically stronger. If two arguments are contextually equally strong, then we refine this ordering by comparing their sets of attack points.

Definition 5 [**Dialectical strength**] Let $AF = (\mathcal{A}, \mathcal{C})$ and $AF' = (\mathcal{A}', \mathcal{C}')$ be two abstract argumentation frameworks in a given UAF and let $A \in \mathcal{A}$ and $B \in \mathcal{A}'$ where the contextual status of A in AF is s and the contextual status of B in AF' is s'. We say that

$A_{AF} \geq_c B_{AF'}$ iff either $s = in$, or $s = und$ and $s' \neq in$, or $s = s' = out$. Moreover, we say that $A_{AF} \geq_d B_{AF'}$ iff

1. $A_{AF} \geq_c B_{AF'}$; and
2. if $B_{AF'} \geq_c A_{AF}$ then $|ap_{AF}(A)| \leq |ap_{AF'}(B)|$.

Below we will leave the subscripts of the arguments implicit if there is no danger of confusion. As usual, $B \leq A$ stands for $A \geq B$ while $A > B$ stands for $A \geq B$ and $B \not\geq A$, and $A \approx B$ stands for $A \geq B$ and $B \geq A$.

The condition that an attack point of A is relevant to A is to exclude examples like an AF with unattacked A and B which are both attacked by the same argument C from UAF: without the relevance condition and if expanding AF with C is allowed, then $\{(B, t)\}$ would (for a given t) be an attack point of A, which is undesirable.

We now illustrate the definition with the AFs in Figure 3. Many current gradual

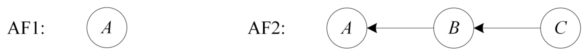

Figure 3. The reinstatement pattern

accounts regard A_{AF1} as stronger than A_{AF2} based on the intuition that having no attackers is better than having attackers (the principle of Void Precedence discussed in the next section). In our approach, this depends on several things. Suppose first that all of A, B and C have attackers in UAF, that A and C have one attack target, respectively, t and t', that A has other unattacked attackers in UAF besides B, and that all expansions are allowed. Then A_{AF1} has just one attack point, namely, $\{(A, t)\}$, while A_{AF2} has two attack points, namely, $\{(A, t), (C, t')\}$. So in this case having no attackers is better.

However, assume now that A has no other attackers in UAF besides B, or that A does have other attackers in UAF but that no expansion with these other attackers is allowed, perhaps for efficiency reasons. In both cases A_{AF1} still has the single attack point $\{(A, t)\}$ but A_{AF2} now also has just one attack point, namely, $\{(C, t')\}$. So here having no attackers is not better than having attackers.

Finally, suppose we change this variation by letting C have no attackers in UAF. Then A_{AF2} has no attack points, so we have a case where an argument that has attackers in one AF is better than an argument that has no attackers in another AF. In conclusion, whether having no attackers is better than having attackers depends on the nature of the arguments and their relations and on the context in which they are evaluated.

4. Properties of Dialectical Argument Strength

We now investigate some properties of our definition of dialectical argument strength. First, $\geq_d$ is a total preorder, that is, transitive and reflexive.

Proposition 1 For all arguments A, B, C and argumentation frameworks AF, AF' and AF'' in a given UAF:

1. $A_{AF} \leq_d A_{AF}$
2. If $A_{AF} \leq_d B_{AF'}$ and $B_{AF'} \leq_d C_{AF''}$ then $A_{AF} \leq_d C_{AF''}$

PROOF. (Sketch:) (1) is immediate, while (2) follows from the facts that both $\leq_c$ and a cardinality ordering on sets are total and that if $A_{AF} \approx_c B_{AF'}$, then a further comparison is made in terms of the cardinality of sets. QED

Definition 6 A UAF satisfies the *attack property* iff for all arguments A, B and C in UAF and all attack targets t that are shared by A and B it holds that C attacks A on t iff C attacks B on t.

The attack property is, for instance, satisfied by assumption-based argumentation in general and by *ASPIC*$^+$ for the case with so-called reasonable argument orderings.

Proposition 2 Consider any UAF satisfying the attack property and let AF be an argumentation framework in UAF containing arguments A and B. Then if $t(A) \subseteq t(B)$ then $A_{AF} \geq_d B_{AF}$.

PROOF. Suppose for contradiction that $A <_d B$ and suppose first that $A <_c B$. If B is *in* but A is not *in* then there exists an attacker C of A that is not *out*. But then C also attacks B so B is not *in*. If B is undecided and A is *out* then there exists an attacker C of A that is *in*. But then C also attacks B so B is *out*. Contradiction.

Suppose next that $A \approx_c B$ and suppose for contradiction that there exists an attack point of A that is not also an attack point of B. This implies that A is not *out* in AF.

Suppose first that A is *in*. Then there exists an allowed expansion AF' of AF where some attackers of A in AF are not *out* and which make that A is not *in* in AF'. By the attack property, these attackers are also attackers of B, so B is not *in* in AF', so the attack point of A is also an attack point of B. Contradiction.

Suppose next that A is *und*. Then there exists an allowed expansion AF' of AF were some attackers of A are *in* and make that A is *out* in AF'. By the attack property, these attackers are also attackers of B, so B is *out* in AF', so the attack point of A also is an attack point of B. Contradiction.

QED

Proposition 2 is what one would expect from dialectical strength as degree of attackability. Its more general version where A and B can be from different AFs does not hold. A counterexample is displayed in Figure 4. Here $\{(A, t)\}$ is (for a given t) an attack point

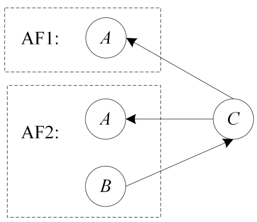

Figure 4. Counterexample to general version of Proposition 2

of A in AF_1 but not in AF_2 since B protects A in AF_2 against an expansion with C. This illustrates that for dialectical strength the dynamic context is important.

Technically our proposal is in the class of ranking-based semantics. We therefore next investigate principles proposed in the literature on ranking-based semantics, basing ourselves on [5]. However, we should first discuss the possible objection that these principles were never intended for dialectical strength, so that investigating them would for present purposes be irrelevant. Against this, it should first be noted that authors are generally not explicit about the kind of strength for which their principles are intended. Moreover, some principles compare different AFs, just as our notion of dialectical strength does, so their underlying intuitions might involve dialectical elements. For these reasons it still makes sense to investigate whether the principles proposed in the literature are suitable for notions of dialectical argument strength. For cases where the underlying intuitions of the proposed principles are not made explicit, our investigation will reveal to which extent they can be based on intuitions concerning dialectical strength.

For reasons of space we have to present the principles discussed in [5] semiformally and we cannot (fully) discuss all of them. When giving counterexamples, we can assume that all considered expansions are allowed.

Proposition 3 Of all principles discussed by [5], Definition 5 only satisfies Attack vs Full Defense and Total.

PROOF. **Total** says that $\leq_d$ is a total ordering. This is stated by Proposition 1. **Attack vs Full Defense** says for acyclic AFs that an argument without any attack branch is ranked higher than an argument only attacked by one non-attacked argument. This holds since any argument of the former kind is *in* while any argument of the latter kind is *out*. QED

For reasons of space we can only give counterexamples to some of the other properties.

Abstraction says that different AFs of the same form should evaluate arguments having the same structural relations in the AFs equally. For a counterexample, consider AF_1 with just A and having one attack target and AF_2 with just B and having two attack targets, where UAF contains additional arguments making that all three attack targets induce the corresponding singleton set attack point. Abstraction says that A and B are of the same rank but we have $A >_d B$. Even if all arguments have the same number of attack targets, there are counterexamples. Assume that both A and B have one attack target and that UAF contains an attacker of B but not of A. Then we again have $A >_d B$.

Void precedence says that a non-attacked argument is ranked strictly higher than any attacked argument in the same AF. One counterexample was given Section 3. Another counterexample is figure 5,which depicts a UAF with an AF in UAF contained

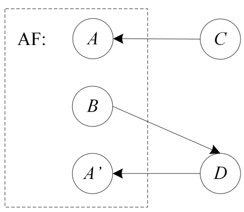

Figure 5. Counterexample to Void Precedence

in the dotted box. Assume all arguments have a single attack target. Then A has attack point $\{(A, t)\}$ but A' has no attack points, since B protects A' against expanding AF with D attacking A.

For two principles that do not hold in general we have identified a special case in which they hold. The **Quality Preference** principle says that if there exists an attacker C of B such that for all attackers D of A it holds that $C >_d D$, then $A >_d B$. It holds in the following special case, since then we have $A >_d B$ so $A >_c B$.

Proposition 4 Def. 5 satisfies Quality Preference if $C >_d D$ for all D attacking A.

A weak version holds of Defence Precedence, which we call **Weak Defense Precedence**, saying that if A_{AF} and B_{AF} have the same number of attackers in AF but A_{AF} has direct defenders while B_{AF} has no direct defender, then $A_{AF} \not<_c B_{AF}$. This holds since an attacked argument with no defenders is always *out*.

Proposition 5 Definition 5 satisfies Weak Defence Precedence.

Why do most principles fail to hold? This is for two main reasons. They fail since they just consider the topology of an AF while dialectical strength also depends on the dynamic context in which an AF can evolve, and/or they fail since they make implicit assumptions on the nature of arguments and their relations that do not hold in general, such as that all arguments have an equal number of attack targets.

5. Dialectical Argument Strength: semantics for weighted AFs

We next adapt our approach to so-called weighted argumentation frameworks [1].

Definition 7 A *weighted argumentation framework* wAF is a triple $(\mathcal{A}_{wAF}, w_{wAF}, \mathcal{C}_{wAF})$ where $\mathcal{A}_{wAF}$ and $\mathcal{C}_{wAF}$ are defined as for AFs and w_{wAF} is a function from $\mathcal{A}_{wAF}$ into $[0, 1]$. A *semantics* for a wAF is another function s_{wAF} from $\mathcal{A}$ into $[0, 1]$.

As above, we omit the subscripts if they are clear from the context. Now a *weighted universal argumentation framework* $wUAF$ is a triple $(\mathcal{A}^u, w^u, \mathcal{C}^u)$, and an AF in UAF is any $wAF = (\mathcal{A}, w, \mathcal{C})$ such that $\mathcal{A} \subseteq \mathcal{A}^u$ and $w = w^u_{|\mathcal{A}}$ and $\mathcal{C} \subseteq \mathcal{C}^u_{|\mathcal{A}\times\mathcal{A}}$. The other of the above definitions for AFs then also apply to $wAFs$ by ignoring w.

We define an argument's weight in a $wUAF$ in terms of its number of attack targets:

$$w_{wUAF}(A) = \frac{1}{1+ \mid t(A) \mid}$$

Note that all weights are between 0 and 1 and that an argument without attack targets has weight 1. We next redefine dialectical argument strength for $wAFs$ as follows.

Definition 8 **[Dialectical argument strength with weights]** An argument's *attack point degree* is defined as $d_{wAF}(A) = \frac{1}{1+|ap_{wAF}A|}$. Then $s_{wAF}(A)$ is defined as follows:

- if A is *in* then $s_{wAF}(A) = \frac{d_{wAF}(A)}{2} + 0.5$;
- if A is *und* then $s_{wAF}(A) = \frac{d_{wAF}(A)}{2}$;
- if A is *out* then $s_{wAF}(A) = 0$.

It can be shown that Definition 8 induces the same ranking on arguments as Definition 5.

Lemma 6 For any wAF and any argument $A \in wAF$ it holds that A is *in* iff $s_{wAF}(a) > 0.5$; A is *und* iff $0 < s_{wAF}(a) \leq 0.5$; and A is *out* iff $s_{wAF}(a) = 0$.

Lemma 7 Let AF and AF' be equal to wAF and wAF' but without weight functions. Then if $A_{AF} \approx_c B_{AF'}$ then $|ap_{AF}(A)| \leq |ap_{AF'}(B)|$ iff $s_{wAF}(B) \leq s_{wAF'}(A)$.

Proposition 8 Let wAF and wAF' be $wAFs$, $A \in wAF$ and $B \in wAF'$ and let AF and AF' be equal to w_{wAF} and $w_{wAF'}$ but without weight functions. Then $A \leq_d B$ iff $s_{wAF}(A) \leq s_{wAF'}(B)$.

PROOF. For the only-if part assume $A \leq_d B$. Two cases must be considered. If $A <_c B$ then $s_{wAF}(A) < s_{wAF'}(B)$ by Lemma 6. If $A \approx_c B$ then $|ap_{AF}(B)| \leq |ap_{AF}(A)|$ so $s_{wAF}(A) \leq s_{wAF'}(B)$ by Lemma 7.

For the if-part assume $s_{wAF}(A) \leq s_{wAF'}(B)$. Two cases must be considered. If $A <_c B$ then $A <_d B$ so $A \leq_d B$. If $A \approx_c B$ then by Lemma 7 we have $|ap_{AF}(B)| \leq |ap_{AF'}(A)|$ so $A \leq_d B$. QED

We next investigate the principles proposed in the literature for semantics of weighted AFs, basing ourselves on [1]. As for ranking-based semantics, for space limitations we cannot discuss all principles while their presentation has to be semiformal.

Proposition 9 Of all principles discussed by [1], Definition 8 only satisfies Weakening Soundness and Compensation.

PROOF. **Weakening soundness** says that for any wAF and any $A \in wAF$, if $s_{wAF}(A) < w_{wAF}(A)$ then there exists an attacker $B \in wAF$ such that $s_{wAF}(B) > 0$. We prove this by contraposition. If there exists no such B, then all attackers of A are *out*. But then A is *in* in wAF. Then $|ap_{wAF}(A)| \leq |t(A)|$, so $d_{wAF}(A) \not< w_{wAF}(A)$. But then $s_{wAF}(A) \not< w_{wAF}(A)$.

Compensation says that there exist wAF in which more weak attackers compensate for fewer stronger attackers. The proof has to specify just one such wAF. Figure 6

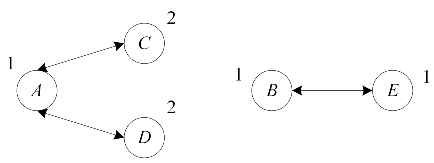

Figure 6. Proof of Compensation

displays a wAF with the number of attack targets of each argument indicated. Assume that all attack targets are an attack point since they have an attacker in UAF (not shown). Note that all arguments are *und*. Then $s_{wAF}(C) = s_{wAF}(D) = \frac{1}{6}$ and $s_{wAF}(E) = \frac{1}{4}$. So A has more attackers with nonzero strength than B while B has an attacker that is stronger than all attackers of A. Moreover, $s_{wAF}(A) = s_{wAF}(B) = \frac{1}{4} > 0$. QED

Counterexamples to the other principles can be constructed as for Definition 5 by considering the context of wAF as defined by $wUAF$ or by considering arguments with sets of attack points of different cardinality. Consider **Monotony**, which says that, for any

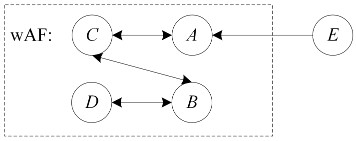

Figure 7. Counterexample to Monotony

$A, B \in wAF$, if $w_{wAF}(A) = w_{wAF}(B)$ and all attackers of A in wAF are attackers of B in wAF, then $s_{wAF}(A) \geq s_{wAF}(B)$. A counterexample is displayed in Figure 7. Here $\{(\mathcal{A}, t)\}$ is an attack point of A since expanding AF with E makes A *out* but no expansion makes B *out*. However, monotony does hold for a special case:

Proposition 10 Monotony holds if all attackers of A in $wUAF$ are attackers of B in $wUAF$ and $wUAF$ satisfies the attack property.

6. Related Research

We do not know of earlier formal work that explicitly addresses dialectical argument strength. Arguably, work on enforcing, preserving or realising a particular argument status [3,8,4], does so implicitly. Compared to this work, we are interested in how the acceptability status of an argument can decrease. A recent structured approach in *ASPIC*$^+$ is [12] (abstracted to AFs in [11]), who study whether argument and conclusion statuses can change under expansions of the knowledge base, to find out whether searching for further information makes sense. It would be interesting to investigate how all this work on argument dynamics can be combined with studies of dialectical argument strength.

As noted above, most work on gradual acceptability does not indicate which aspect(s) of argument strength is or are modelled. A recent exception is [6], who model two aspects of "persuasiveness',' i.e., of rhetorical strength. The first is *procatalepsis*, the attempt of a speaker to strengthen their argument by dealing with possible counterarguments before the audience can raise them. The second aspect is *fading*, the phenomenon that long lines of argumentation are less persuasive. Bonzon et al. claim that "current ranking-based semantics are poorly equipped to be used in a context of persuasion". Among other things, they show that procatalepsis violates the Void Precedence principle. While we agree with their observation, we note that in the end they do not give a separate model of persuasiveness but combine these two aspects with existing strength principles into an overall measure of argument strength, thereby still conflating the three kinds of argument strength. We instead prefer to separately study different notions of argument strength, since these notions may serve different purposes and may therefore evaluate the same arguments differently.

7. Conclusion

In this paper we presented the first formal study of dialectical argument strength, modelled as the number of ways in which an argument can be successfully attacked in expansions of an argumentation framework. We showed that most principles for gradual argument acceptability proposed in the literature fail to hold for the new notion, which reveals something about the possible rational foundations of these principles and highlights the importance of distinguishing between kinds of argument strength. Our model is abstract but its design is motivated by the wish to avoid overly limiting assumptions on dialogue contexts or the structure of arguments and their relations.

Are our partly negative results on satisfaction of the principles bad for our approach or for the principles? There is no easy answer to this question but we note that in the literature most principles are based on intuitions instead of on philosophical insights. Therefore it is not obvious why they should hold; it may just as well be that if a semantics based on philosophical insights and arguably reflecting good properties does not satisfy some principle, then this indicates that the principle may not be suitable for the modelled notion. Our semantics is based on [15] and arguably satisfies desirable properties. In particular, we believe that Proposition 2 and the satisfaction of Weakening Soundness and the special case of Monotony indicate that our semantics captures the ideas of [15] and the intuition that justifying a decision more sparsely is better.

In future research we want to extend our abstract model with support relations between arguments and to study structured instantiations of our model and applications to particular dialogue contexts. We also want to extend our approach to semantics other than grounded semantics, including gradual semantics.

References

[1] L. Amgoud, D. Doder, and S. Vesic. Evaluation of argument strength in attack graphs: foundations and semantics. *Artificial Intelligence*, 302:103607, 2022.

[2] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26:365–410, 2011.

[3] R. Baumann and G. Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In *Proceedings of COMMA 2010*, pages 75–86, 2010.

[4] D. Baumeister, M. Järvisalo, D. Neugebauer, A. Niskanen, and J. Rothe. Acceptance in incomplete argumentation frameworks. *Artificial Intelligence*, 295:103470, 2021.

[5] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. A comparative study of ranking-based semantics for abstract argumentation. In *Proceedings of AAAI 2016*, pages 914–920, 2016.

[6] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. A parametrized ranking-based semantics compatible with persuasion principles. *Argument and Computation*, 12:49–85, 2021.

[7] 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.

[8] S. Doutre and J.-G. Mailly. Constraints and changes: A survey of abstract argumentation dynamics. *Argument and Computation*, 9:223–248, 2018.

[9] 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.

[10] A.J. Hunter, editor. *Argument and Computation*, volume 5. 2014. Special issue with Tutorials on Structured Argumentation.

[11] J.-G. Mailly and J. Rossit. Stability in abstract argumentation. In *Proceedings of the 18th International Workshop on Nonmonotonic Reasoning*, pages 93–99, 2020.

[12] D. Odekerken, A. M. Borg, and F.J. Bex. Estimating stability for efficient argument-based inquiry. In *Proceedings of COMMA 2020*, pages 307–318, 2020.

[13] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.

[14] H. Prakken. Philosophical reflections on argument strength and gradual acceptability. In *Proceedings of ECSQARU 2021*, pages 144–158, 2021.

[15] F. Zenker, K. Debowska-Kozlowska, D. Godden, M. Selinger, and S. Wells. Five approaches to argument strength: probabilistic, dialectical, structural, empirical, and computational. In *Proceedings of the 3rd European Conference on Argumentation*, pages 653–674, London, 2020. College Publications.