An Abstract and Structured Account of Dialectical Argument Strength

Henry Prakken[1]
1. Department of Information and Computing Sciences, Utrecht University, The Netherlands
(H.Prakken@uu.nl)

July 30, 2024

Abstract

This paper presents a formal model 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. First a model is proposed that is abstract but designed to avoid overly limiting assumptions on instantiations or dialogue contexts. It is then shown that most principles for argument strength proposed in the literature fail to hold for the proposed notions of dialectical strength, which clarifies the rational foundations of these principles and highlights the importance of distinguishing between kinds of argument strength, in particular logical, dialectical and rhetorical argument strength. The abstract model is then instantiated with $ASPIC^+$ to test the claim that it does not make overly limiting assumptions on the structure of arguments and the nature of their relations.

Keywords: Computational argumentation, Expansion games, Dialectical argument strength, Structured argumentation frameworks.

1 Introduction

Argumentation is a key topic in the logical study of nonmonotonic reasoning and the dialogical study of inter-agent communication [7], and has received much attention from the Artificial Intelligence (AI) community since the late 1980s. Argumentation as a form of reasoning makes explicit the reasons for the conclusions that are drawn and how conflicts between reasons are resolved. This provides a natural mechanism to handle inconsistent and uncertain information and to resolve conflicts of opinion between intelligent agents. In logical models of nonmonotonic reasoning, the argumentation metaphor arguably overcomes some drawbacks of other formalisms. Many of these have a mathematical nature that is remote from how people actually reason, which makes it difficult to understand and trust the behaviour of an intelligent system. The argumentation approach aims to bridge this gap by providing logical formalisms that are rigid enough to be formally studied and implemented, while at the same time being close enough to informal reasoning to be understood by designers and users [49]. This makes the formal study of argumentation very relevant to current research in explainable AI, which is the subfield of AI that studies how the behaviour and results of, often non-transparent, AI algorithms can be explained to humans [46, 47, 72]. Formal models of argumentation are also relevant as benchmarks of argumentative applications of generative AI [67, 37].

A recent trend in the formal study of argumentation is the development of gradual notions of argument acceptability or argument strength. These notions are proposed as alternatives to extension-based notions that are defined on top of the theory of abstract [25] or bipolar [22] argumentation frameworks. The gradual notions are often motivated by a discontent with the fact that extension-based notions of acceptability only allow for rather coarse distinctions between degrees of acceptability, which would not fit with the more nuanced ways in which humans evaluate arguments. Pollock [57] was, to our knowledge, the first who addressed this issue and proposed a formalisation of gradual "justification". The current developments arguably go back to [21] and gained momentum with publications like [43] and [2]. The current studies include probabilistic [41], gradual [33] and ranking-based [2] approaches.

Although the new developments are very interesting and the formal achievements have been impressive, there are also reasons to take a step back. To start with, there is a need to reflect on which notions or aspects of argument acceptability, or argument strength, are modelled, and why proposed semantics or proposed sets of principles for those semantics are good. What is needed is a conceptual and philosophical underpinning of the formal ideas and constructs. Furthermore, almost all work builds on abstract or bipolar argumentation frameworks and thus does not give explicit formal accounts of the nature of arguments and their relations, while yet this may be relevant when evaluating the formal proposals. This paper[1] addresses both issues.

Before doing so, a remark on terminology is in order. The use of the terms 'strength' and 'acceptability' in the literature varies. While e.g. [11, 4, 33] use these terms interchangeably, Amgoud [1] proposes that strength and acceptability are different concepts in that acceptability is about which arguments can be (jointly) accepted by an agent, where argument strength is one aspect that may determine argument acceptability. It seems to us that Amgoud's proposal is only about what we will below call contextual argument strength, while we will distinguish several kinds of argument strength. Because of this and since there is no consensus in the literature yet about the use of these terms, we will use 'strength' throughout the paper and leave open the possibility that for contextual strength this is interpreted as degree of acceptability in the sense of [1].

1.1 Kinds of Argument Strength

As for the first issue mentioned above, which notions of argument strength are modelled, we argue that work on gradual argument strength should make explicit which kind of argument strength is modelled, since different kinds of strength may be subject to different rationality constraints. In this paper we take Aristotle's famous distinction between logic, dialectic and rhetoric as starting point. Very briefly, *logic* concerns the validity of arguments given their form, *dialectic* is the art of testing ideas through critical discussion and *rhetoric* deals with the principles of effective persuasion [71, Section 1.4]. Accordingly, we distinguish between logical, dialectical and rhetorical argument strength, where logical argument strength in turn divides into two aspects: inferential and contextual argument strength.

[1]This paper combines, extends and generalises [61, 62, 63]. In particular, Section 1.1 is adapted from [61], Section 3 generalises the 'single-shot' expansion approach of [62] to a setting with expansion games, Section 4 extends and generalises the results of [62] for one generalised and one new contextual argument ordering, while Sections 5.1 and 5.2 are based on definitions of [63]. Section 5 is fully new while Section 7 combines discussions of [62, 63] with new discussions.

Inferential argument strength is about how well an argument's premises support its conclusion considering only the argument itself. Example criteria for argument strength are that arguments with only deductive inferences are stronger than arguments with defeasible inferences, or that arguments with only non-attackable premises are stronger than arguments with attackable premises.

Contextual argument strength is about how well the conclusion of an argument is supported in the context of a given set of 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 [42]. Arguably some recent ranking and gradual semantics also aim to model contextual strength, witness e.g. the following quote from [1]:

> [An argument's] strength depends on the plausibility of the premises, the strength of the link between the premises and claim, and the prior acceptability of the claim. Attacks aim to highlight weaknesses in these three components of an argument. Hence, the less an argument is attacked, the stronger it is.

The reader might wonder why contextual strength is not called dialectical strength, since after all, determining an argument's contextual strength as defined here involves the comparison of arguments and counterarguments. Yet this is not truly dialectical, since the just-mentioned formalisms do not model principles of critical discussion but define structural relations between (sets of) arguments on the basis of a given body of information; likewise [30, 44].

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 (as acknowledged in [40] and done in e.g. [34]).

Dialectical argument strength looks at how challengeable an argument is in the context of a critical discussion. In [77, 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 vulnerability of an argument in that how many attacks are allowed in a given state that decrease the argument's contextual strength. 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.

A separate study of dialectical argument strength is justified since the three aspects of argument strength serve different purposes, so it may not be good to combine them into an overall notion. Another reason for this is that dialectical strength *presupposes* contextual strength, since one aspect of dialectical strength is the extent to which an argument's contextual strength may be changed in the course of a dispute. In any case, even if logical, dialectical and rhetorical strength are combined into an overall notion of strength, they should first be separately defined, in order to make their combination a principled one.

Being explicit about which aspects of argument strength are modelled is not only important when formulating theories of argument strength but also when evaluating applications of computational argumentation. For instance, according to [67] the Debater system was evaluated by twenty human annotators, who had to indicate to what extent they agreed with the statement 'The first speaker is exemplifying a decent performance in this debate'. It is unclear which aspects the annotators had in mind when answering this question or even whether all annotators looked at the same aspects and applied the same criteria. With the present study, we aim to contribute to more principled methods for evaluating argumentation tools like Debater.

1.2 Relating Abstract and Structured Approaches

To address the second issue mentioned above, relating abstract and structured approaches, we will first propose an abstract formal model 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, and we will then instantiate it with the $ASPIC^+$ framework [50, 51]. The choice for $ASPIC^+$ is motivated by the facts that it is well-studied and often applied (e.g. [10, 28, 60, 52, 53, 66, 68, 70]) while variants of assumption-based [69] and classical [31] argumentation can be reconstructed as special cases of $ASPIC^+$ [51].

The abstract model is defined in terms of a refined version of the notion of a normal expansion of an abstract argumentation framework as proposed in [8], Although the model is abstract, its design is motivated by the wish to avoid overly limiting assumptions on instantiations or dialogue contexts. Illustrating its adequacy in this respect is one aim of the instantiation with $ASPIC^+$. Another aim of the instantiation is to study which properties that do not hold in general may hold if assumption are made on the structure of arguments and the nature of their relations. Among other things, we will show that most principles for argument strength proposed in the literature fail to hold in general for this paper's notion of dialectical strength, both for its abstract version and its instantiation with $ASPIC^+$. We will argue that this casts doubt on the rational foundations of these principles.

1.3 Overview of the Paper

The rest of paper is organised as follows. In Section 2 we summarise the formalisms used in this paper: the theory of abstract argumentation frameworks and the $ASPIC^+$ framework. In Section 3 we present and study our abstract model of dialectical strength in ranking-based form, which in Section 4 we alternatively combine with two semantics for contextual strength: the standard semantics based on [25] and the ranking-based burden semantics of [2]. We then instantiate our abstract approach with $ASPIC^+$ in Section 5 and we investigate the formal properties of this instantiation in Section 6. In Section 7 we discuss related work and use it to make some preliminary observations on computational complexity. We conclude in Section 8 with a discussion of what we have achieved and what is left to be done, and of the wider implications of our results for AI research on models of argument strength.

2 Formal Preliminaries

In this section we summarise the formal preliminaries: the theory of abstract argumentation frameworks and their expansions, and the *ASPIC*$^+$ framework.

2.1 Abstract Argumentation Frameworks: Semantics and Expansions

An *abstract argumentation framework* [25] is a pair $AF = (\mathcal{A}_{AF}, \mathcal{D}_{AF})$, where $\mathcal{A}_{AF}$ is a set of arguments and $\mathcal{D}_{AF} \subseteq \mathcal{A}_{AF} \times \mathcal{A}_{AF}$ is a relation of defeat.[2] 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 $\mathcal{D} = \{(B, A), (B, C), (C, B)\}$. Let $S \subseteq A$. Then S is conflict-free if no member of S defeats a member of S and S *defends* $A \in \mathcal{A}$ if for all $B \in \mathcal{A}$: if B defeats A, then some $C \in S$ defeats B.

Then relative to a given AF, the following semantics were defined in [25], which we will call *classical semantics* for abstract argumentation framework.

Definition 1 [Classical semantics for AFs] Let $(\mathcal{A}_{AF}, \mathcal{D}_{AF})$ be an AF and $E, S \subseteq \mathcal{A}_{AF}$. Then

- E 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 defeats all arguments outside it;
- $E \subseteq \mathcal{A}$ 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. For $x \in$ {complete, preferred, grounded, stable}[3], X is *skeptically* justified under the x semantics if X belongs to all x extensions. For a notion of credulous justification we distinguish between grounded semantics and the three other semantics. For $x \in$ {complete, preferred, stable} X is *credulously* justified under the x semantics if X belongs to at least one x extension. Under grounded semantics, X is *credulously* justified if X does not belong to the grounded extension but is not defeated by an argument in the grounded extension.

In this paper we will often use an equivalent labelling way to define semantics for AFs. A *labelling* of a set $\mathcal{A}$ of a set of arguments in an $AF = (\mathcal{A}, \mathcal{D})$ is any partitioning (*in*,*out*,*und*) of $\mathcal{A}$ that satisfies the following constraints:

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

[2] Dung used the term 'attack' but since we want to instantiate it with the *ASPIC*$^+$ defeat relation, we rename it to 'defeat'.

[3] In later papers new semantics have been introduced, see Baroni, Caminada, and Giacomin [6], but we only discuss these semantics of [25].

Here *und* stands for 'undecided'. Then *stable semantics* labels all arguments, while *grounded semantics* minimises and *preferred semantics* maximises the set of arguments that are labelled *in* and *complete semantics* allows any labelling. It has been shown for $x \in$ {complete, preferred, grounded, stable} that the set of x-extensions is equal to the set of all sets S such that S is a set of all arguments labelled *in* in a particular x-labelling [18].

Ranking-based semantics for abstract argumentation frameworks are defined as follows [11]:

Definition 2 [Ranking-based semantics for AFs] A *ranking-based semantics* for an $AF = (\mathcal{A}_{AF}, \mathcal{D}_{AF})$ is a preorder $\geq_{AF}$ on $\mathcal{A}_{AF}$, that is, a transitive and reflexive ordering. That $A \geq_{AF} B$ means that A is at least as acceptable as B (Henceforth, the subscript AF will be omitted if there is no danger for confusion.) As usual, $A > B$ is defined as $A \geq B$ and $B \not\geq A$ and $A \approx B$ as $A \geq B$ and $B \geq A$.

In our paper we will define comparisons between arguments that possibly are in different AFs. Accordingly, we will later generalise the notion of ranking-based semantics as applying to sets of argument-framework pairs.

The following definition is from [11], adapted to our notation and terminology, and defines several notions used in postulates proposed in the literature for ranking-based semantics.

Definition 3 Let $AF = (\mathcal{A}, \mathcal{D})$ and $A, B \in \mathcal{A}$. A *path* P from B to A, denoted $P(B, A)$, is a sequence $S = (A_0, \ldots, A_n)$ of arguments such that $A_0 = A$, $A_n = B$ and $\forall i < n, (A_{i+1}, A_i) \in \mathcal{D}$. We denote by $l_P = n$ the *length* of P. A *defender* (resp. *defeater*) of A is an argument situated at the beginning of an even-length (resp. odd length) path ending with A. We denote the multiset of defenders and defeaters of A by $R_n^+(A) = \{B \mid \exists P(B, A)$ with $l_P \in 2\mathbb{N}\}$ and $R_n^-(A) = \{B \mid \exists P(B, A)$ with $l_P \in 2\mathbb{N} + 1\}$ respectively[4]. An argument A is *defended* if $R_2^+(A) \neq \emptyset$.

The *direct defeaters*, (resp. *direct defenders*) of A are the arguments in $R_1^-(A)$ (resp $R_2^+(A)$. Below we will simplify the notations $R_1^-(A)$, respectively, $R_2^+(A)$ to A^-, respectively, A^+. Also, henceforth we will use the term 'defeater' for direct defeaters and call all other defeaters 'indirect defeaters'.

A *defense root* (resp. *defeat root*) is a non-defeated defender (resp. direct or indirect defeater). We denote the multiset of defense roots and defeat roots of argument A by $BR_n^+(A) = \{B \in R_n^+(A) \mid |B^-| = 0\}$ and $BR_n^-(A) = \{B \in R_n^-(A) \mid |B^-| = 0\}$ respectively. A path from B to A is a *defense branch* (resp. *defeat branch*) if B is a defense (resp. defeat) root of A. Let us denote $BR^+(A) = \bigcup_n BR_n^+(A)$ and $BR^-(A) = \bigcup_n BR_n^-(A)$.

Finally, Baumann and Brewka [8] define various kinds of expansions of abstract argumentation frameworks as follows.

Definition 4 **[Expansions]** An abstract argumentation framework AF' is an *expansion* of an abstract argumentation framework $AF = (A, \mathcal{D})$ iff $AF' = (\mathcal{A} \cup \mathcal{A}', \mathcal{D} \cup \mathcal{D}')$ for some nonempty $\mathcal{A}'$ disjoint from $\mathcal{A}$. An expansion is

1. *normal* iff for all A, B: if $(A, B) \in \mathcal{D}'$ then $A \in \mathcal{A}'$ or $B \in \mathcal{A}'$,
2. *strong* iff it is normal and for all A, B: if $(A, B) \in \mathcal{D}'$ then it is not the case that $A \in \mathcal{A}$ and $B \in \mathcal{A}'$,

[4] $2\mathbb{N}$ denotes the set of all even natural numbers.

3. *weak* iff it is normal and for all A, B: if $(A, B) \in \mathcal{D}'$ then it is not the case that $A \in \mathcal{A}'$ and $B \in \mathcal{A}$.

Normal expansions add new arguments and possibly new defeat relations, where new defeats should involve at least one new argument. Strong, respectively, weak expansions are normal expansions that do not add defeats from old to new, respectively, from new to old arguments.

2.2 The *ASPIC*$^+$ Framework

In this section we summarise the *ASPIC*$^+$ framework for strructured argumentation. Over the years several variants of the *ASPIC*$^+$ framework have been developed and studied. In this paper we use a special case of the 'basic' framework as formulated in [59] and further studied in [49]. The special case is that we consider a language with symmetric negation, noting that all new definitions proposed in Section 5 of this paper can be easily adapted to the versions of *ASPIC*$^+$ with asymmetric negation, while the counterexamples given in Section 6 of this paper directly hold for these generalisations. On the other hand, we generalise the basic system in that we do not consider specific ways to define the preference relation on arguments. At the end of this section we will briefly discuss other variants of *ASPIC*$^+$ that have been proposed.

2.2.1 Basic Definitions

ASPIC$^+$ defines abstract argumentation systems as structures consisting of a logical language $\mathcal{L}$ and two sets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules defined over $\mathcal{L}$. Arguments are constructed from a knowledge base (a subset of $\mathcal{L}$) by chaining inferences over $\mathcal{L}$ into acyclic graphs (which are trees if no premise is used more than once). Formally,

Definition 5 [**Argumentation System**] an *argumentation system* (AS) is a triple $AS = (\mathcal{L}, \mathcal{R}, n)$ where:

- $\mathcal{L}$ is a logical language with a negation symbol $\neg$;
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a 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 well-formed formulas (wff) in $\mathcal{L}$), such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$. Here, $\varphi_1, \ldots, \varphi_n$ are called the *antecedents* and φ the *consequent* of the rule.
- n is a partial function such that $n : \mathcal{R}_d \longrightarrow \mathcal{L}$.

Informally, $n(r)$ is a wff in $\mathcal{L}$ which says that the defeasible rule $r \in \mathcal{R}$ is applicable, so that an argument claiming $\neg n(r)$ attacks an inference step in the argument using r. We write $\psi = -\varphi$ just in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$. We use $\rightsquigarrow$ as a variable ranging over $\{\rightarrow, \Rightarrow\}$. Since the order of antecedents of a rule does not matter, we sometimes write $S \rightsquigarrow \varphi$ where S is the set of all antecedents of the rule.

Definition 6 [**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*).

Definition 7 [**Argumentation theories**] An *argumentation theory* is a pair $(AS, \mathcal{K})$ where AS is an argumentation system and $\mathcal{K}$ a knowledge base in AS.

Definition 8 [**Arguments**] An *argument* A on the basis of an argumentation theory AT is a structure obtainable by applying one or more of the following steps finitely many times:

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Prop}(A) = \{\varphi\}$, $\texttt{Sub}(A) = \{\varphi\}$; $\texttt{Rules}(A) = \emptyset$; $\texttt{DefRules}(A) = \emptyset$; $\texttt{TopRule}(A)$ = undefined.
2. $A_1, \ldots, A_n \rightsquigarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that $\psi \notin \texttt{Conc}(\{A_1, \ldots, A_n\})$ and $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightsquigarrow \psi \in \mathcal{R}$ with:
$\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$;
$\texttt{Conc}(A) = \psi$;
$\texttt{Prop}(A) = \texttt{Prop}(A_1) \cup \ldots \cup \texttt{Prop}(A_n) \cup \{\psi\}$,
$\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$;
$\texttt{Rules}(A) = \texttt{Rules}(A_1) \cup \ldots \cup \texttt{Rules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightsquigarrow \psi\}$;
$\texttt{DefRules}(A) = \texttt{Rules}(A) \cap \mathcal{R}_d$;
$\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightsquigarrow \psi$.

$\texttt{Prem}_n(A) = \texttt{Prem}(A) \cap \mathcal{K}_n$ and $\texttt{Prem}_p(A) = \texttt{Prem}(A) \cap \mathcal{K}_p$. Furthermore, argument A is *strict* if $\texttt{DefRules}(A) = \emptyset$ and *defeasible* otherwise, and A is *firm* if $\texttt{Prem}_p(A) = \emptyset$, otherwise A is *plausible*.

The set of all arguments on the basis of AT is denoted by $\mathcal{A}_{AT}$.

Each of the functions $\texttt{Func}(A)$ in this definition is also defined on sets of arguments $S = \{A_1, \ldots, A_n\}$ as follows: $\texttt{Func}(S) = \texttt{Func}(A_1) \cup \ldots \cup \texttt{Func}(A_n)$. Note that the $\rightarrow$ and $\Rightarrow$ symbols are overloaded to denote both inference rules and arguments.

Definition 9 [**Attack**] Argument A *attacks* argument B iff A *undercuts* or *rebuts* or *undermines* B, where:

- A *undercuts* 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* B (on B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$.
- A *undermines* B (on φ) iff $\texttt{Conc}(A) = -\varphi$ for some $\varphi \in \texttt{Prem}(B) \cap \mathcal{K}_p$.

Definition 10 [**Structured Argumentation Frameworks**] A *structured argumentation framework (SAF)* defined by an argumentation theory AT is a triple $(\mathcal{A}, \mathcal{C}, \preceq)$ where $\mathcal{A}$ is the set of all arguments on the basis of AT, $\preceq$ is an ordering on $\mathcal{A}$ and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

The notion of *defeat* is now 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, where $A \prec B$ is defined as usual as $A \preceq B$ and $B \npreceq A$ and $A \approx B$ as $A \preceq B$ and $B \preceq A$. Below we assume that $\prec$ is asymmetric while, moreover, if A is strict and firm, then $A \prec B$ does not hold.

Definition 11 [**Defeat**] Argument A *defeats* argument B iff either A undercuts B; or A rebuts or undermines B on B' and $A \nprec B'$.

Abstract argumentation frameworks are then generated from $SAFs$ as follows:

Definition 12 [**Argumentation frameworks**] An *abstract argumentation framework (AF) corresponding to an* $SAF = (\mathcal{A}, \mathcal{C}, \preceq)$ is a pair $(\mathcal{A}, \mathcal{D})$ such that $\mathcal{D}$ is the defeat relation on $\mathcal{A}$ determined by SAF.

2.2.2 Research on Consistency Postulates and Variants of *ASPIC*$^+$

In [19] postulates were studied that say that the set of conclusions of all arguments in an extension should be directly and indirectly consistent. Informally, a set of wffs is *directly consistent* if it does not contain a formula and its negation and it is *indirectly consistent* if its closure under strict-rule application is directly consistent. The *ASPIC*$^+$ framework as presented thus far leaves one fully free to choose a language, what is an axiom and what is an ordinary premise, how to specify strict and defeasible rules and how to define the preference relation between arguments. As a consequence, *ASPIC*$^+$ does not in general satisfy these consistency postulates. It has been shown that *ASPIC*$^+$ shares this feature with most other structured approaches to argumentation, such as Defeasible Logic Programming [19], classical-logic argumentation [31] and assumption-based argumentation [49]. For *ASPIC*$^+$ much research has been done on identifying broad and well-behaved classes of instantiations that do satisfy direct and indirect consistency; see e.g. [19, 59, 49, 26, 76, 32]. Since the way we will use *ASPIC*$^+$ in this paper does not depend on satisfaction of these postulates, our approach and results will apply to all these classes of instantiations.

Furthermore, noted above, over the years several variants of the *ASPIC*$^+$ framework have been developed and studied, all with their strengths and weaknesses, and all with new results on satisfaction of the consistency postulates. In [49] in fact four variants are studied, along two axes: whether the premises of arguments have to be indirectly consistent or not, and whether conflict-freeness of sets of arguments is defined in terms of the attack or the defeat relation. In [19, 20, 36] variants of rebutting attack called 'unrestricted rebut' are studied, in which arguments can also be rebutted on conclusions of strict top rules, provided that at least one subargument of the attacked argument is attackable. In [32] a variant is proposed to deal with the so-called contamination problem when $\mathcal{R}_s$ is generated by classical logic. As explained in [17], in such cases rebutting arguments give rise to unwanted further arguments because of the Ex Falso property (a problem that can also arise in assumption-based argumentation). In [32] this problem is avoided by letting $\mathcal{R}_s$ be generated by so-called weak consequence [65], which says that something weakly follows from a set iff it follows classically from at least one consistent subset of that set. This requires that strict rules cannot be chained in arguments, since weak consequence does not satisfy the cut rule. Finally, in [23] a variant of *ASPIC*$^+$ is proposed to deal with contamination problems in a different way, namely, by modifying the definitions of arguments, attacks and defeats in order to allow for expressing whether premises or arguments are committed to or instead supposed for the sake of argument. The resulting system can also model argument evaluation under resource bounds.

In this paper we have chosen to work with the 'basic' variant of [59, 49] since it is the simplest and most widely known and used variant and since the way we will instantiate our abstract account of Sections 3 and 4 does not depend on the particular features of the adopted variant of *ASPIC*$^+$. Accordingly, we use the basic version as a representative example and we conjecture that our approach and results easily extend to the other variants.

3 Dialectical Argument Strength: Abstract Ranking-Based Semantics

In this section we define an abstract ranking-based semantics of dialectical strength of arguments. As noted in Section 2, we will not as usual define it on the set of arguments of a single given AF, but instead on the set of all argument-AF pairs given a set of AFs.

3.1 Ideas

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 [77] 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. First of all, such an approach would need a clear definition of when and how dialogues terminate, but in many contexts (such as informal debate) such a clear definition cannot be given. Moreover, it will often be impossible to foresee which information will become available to construct arguments, how arguments will be evaluated in the course of a dialogue and which procedural decisions (such as on admissibility of evidence or termination of a dialogue) will be taken. And even if all this can be foreseen in theory, there will in practice often be computational or resource limitations. Although there can be specific contexts where considering all possible completions of the current state is feasible, a general account cannot rely on this.

For these reasons, we propose the following flexible approach. We define the extent to which an argument can be successfully attacked in continuations of a dialogue in terms of a dialogue game in which an opponent and proponent of a given focus argument successively expand the current state in a way that, respectively, decreases and increases the 'current' contextual status of the focus argument. Successful attack is then defined in terms of the existence of a winning strategy for the opponent in this game. The flexibility is given by the fact that the notion of a winning strategy is made relative to a given maximal length of a dialogue game. The choice of a suitable maximum length depends on the context and the nature of the application.

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 defeat in all these expansions. To this end we will define a notion of *defeat points*[5] of an argument, which are minimal sets of arguments that, if defeated in an allowed expansion, make the contextual strength of the focus argument decrease. We will formally define this notion in Section 3.4. Let us, to explain the intuitions, first focus on 'single-shot' games, which terminate after the first move by the opponent (the setting of [62]).

Example 1 Consider an $AF = A \leftarrow B$, let A be the focus argument and let arguments

[5] In [62] defeat points were called 'attack points'.

be evaluated using grounded semantics. Assume that 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 at first sight it would seem that it does not matter whether the proponent moves C or D. However, assume that the arguments are generated in *ASPIC*$^+$ and that C can be defeated since it is defeasible while D cannot be defeated since it is strict and firm. Assume, moreover, that the opponent can expand AF' to a framework AF''' in which a defeat of C results in A being *und* or *out*. Then A has two defeat points in AF', namely, $\{A\}$ and $\{C\}$, while A has only one defeat point in AF'', namely, $\{A\}$. So $A_{AF''}$ is dialectically stronger than $A_{AF'}$, therefore, the dialectically better choice for the proponent is to expand AF to AF'' by adding D.

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 defeat points as follows. To start with, we assume a contextual argument ordering for a (possibly infinite) set of AFs. Since in general we want to compare arguments in different AFs, we define it as an ordering on argument-AF pairs. Moreover, since we want to allow for contextual orderings generated by different argumentation semantics, we make it relative to a semantics x.

Definition 13 Given a semantics x for abstract argumentation frameworks, a *contextual ordering of argument strength* for a set S of abstract argumentation frameworks is a preorder $\geq_c^x$ defined on $\{(A, AF) \mid AF \in S \text{ and } A \in AF\}$.

Below, when S and/or x is clear from the context, it will be omitted, and $A_{AF} \geq_c B_{AF'}$ will stand for $(A, AF) \geq_c (B, AF')$, where the subscripts of the arguments will be left implicit if AF and AF' are clear from the context. As usual, $B \leq_c A$ stands for $A \geq_c B$ while $A >_c B$ stands for $A \geq_c B$ and $B \not\geq_c A$, and $A \approx_c B$ stands for $A \geq_c B$ and $B \geq_c A$. In Section 4 several instantiations of x in $\geq_c^x$ will be studied. For now the nature of x will be left fully unspecified in the formal theory to be developed while in examples the following simple contextual ordering for the labelling version of grounded semantics (g) will be used (which was also assumed in [62]): *in* $>_c^g$ *und* $>_c^g$ *out*.

Then, given the set of allowed expansions $\{AF', AF'', \ldots\}$ of a given AF, the idea is to 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 defeat points than $B_{AF''}$, that is, if there are fewer ways to successfully lower the contextual strength of A than that of B (in a way to be made precise below). 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.

3.2 Expansions and Expansion Games

We now define the notions of allowed expansions and an expansion game, which with various other definitions will allow us to define defeat points of an argument.

3.2.1 Expansions in universal argumentation frameworks

To define defeat points, we now first define the notion of an allowed expansion of an AF, which is a refinement of [8]'s notion of an expansion. The first refinement is to make expansions relative to a given background universal argumentation framework $UAF = (\mathcal{A}^u, \mathcal{D}^u)$ from which expansions can take their new arguments and defeats. Arguably any model of expansions needs to fix a universal background, otherwise there is no way to identify possible expansions. Therefore, it is worthwhile to make notions of background information explicit in order to develop a theory about them. Another important reason for doing so is that the notion of a universal background AF helps avoiding implicit assumptions at the abstract level that are not always satisfied by instantiations, such as that all arguments can be defeated or that all defeats are independent from each other. We will come back to this point below.

Definition 14 [**Argumentation frameworks in a universal *AF***] Given a universal argumentation framework $UAF = (\mathcal{A}^u, \mathcal{D}^u)$, an *argumentation framework in UAF* is any $AF = (\mathcal{A}, \mathcal{D})$ such that $\mathcal{A} \subseteq \mathcal{A}^u$ and $\mathcal{D} \subseteq \mathcal{D}^u_{|\mathcal{A}\times\mathcal{A}}$.

The fact that $\mathcal{D}$ is not required to equal $\mathcal{D}^u_{|\mathcal{A}\times\mathcal{A}}$ is to allow for instantiations with systems like *ASPIC*$^+$ that use preferences to resolve attacks. For example, suppose $A, B \in \mathcal{A}^u$ and $(A, B), (B, A) \in \mathcal{D}^u$. Then $AF = (\{A, B\}, \{(A, B)\})$ is an AF in UAF that expresses a preference of A over B. Nevertheless, below we will sometimes consider AFs in a UAF that *do not omit defeats from* a UAF in that $\mathcal{D} = \mathcal{D}^u_{|\mathcal{A}\times\mathcal{A}}$.

We must also distinguish between allowed and unallowed 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 forbidden. 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 [42] may impose such constraints, as will be shown in detail in Section 5 below. For now we give a simple example in *ASPIC*$^+$.

Example 2 Suppose we have an AF consisting of the following arguments:

A_1: p, A_2: $p \Rightarrow q$

where $p \in \mathcal{K}_p$. Suppose *UAF* contains the following defeater of A_1:

B: $\neg p$

where $\neg p \in \mathcal{K}_n$. Then expanding AF to $AF' = (\{A_1, A_2, B\}, \{(B, A_1)\})$ should not be allowed, since in *ASPIC*$^+$ B also defeats A_2. So this 'implied' defeat relation (B, A_2) must als be added to the expansion.

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

Definition 15 [**Expansions given a universal argumentation framework**] Let $AF = (\mathcal{A}, \mathcal{D})$ and $AF' = (\mathcal{A}', \mathcal{D}')$ be two abstract argumentation frameworks in a *UAF*. Then AF' is an *expansion* of AF given *UAF* if $AF' = (\mathcal{A} \cup \mathcal{A}', \mathcal{D} \cup \mathcal{D}')$ for some nonempty $\mathcal{A}'$ disjoint from $\mathcal{A}$. The notions of a normal, weak and strong expansion in *UAF* are defined as the corresponding notions in Definition 4.

Let $X_{UAF}(AF)$ be the set of all expansions of AF given *UAF*. Then the set of *allowed expansions* of AF given *UAF* is some designated subset of $X_{UAF}(AF)$.

3.2.2 Expansion games

Recall that we want to define the extent to which an argument can be successfully attacked in continuations of a dialogue in terms of a dialogue game in which an opponent and proponent of a given focus argument successively expand the current state in a way that, respectively, decreases and increases the 'current' contextual status of the focus argument. Successful attack can then defined in terms of the existence of a winning strategy for the opponent in this game. To formalise these ideas, we must first define an expansion game of length n between a proponent and an opponent of a given argument A in an initial AF. The opponent starts the game by expanding the AF to AF' in such a way that the contextual status of A is reduced in AF'. Then the players take turns. The proponent must extend the last expansion of the opponent in such a way that A's contextual status is increased to at least its original status, while the opponent must extend the last expansion of the proponent in such a way that A's contextual status is reduced to below its original status. A player wins a finite game if the other player has no legal reply.

These ideas are formalised as follows.

Definition 16 [**Expansion game**] Let A be any argument in a AF in a given UAF and let $n \in \mathbb{N} \cup \{\infty\}$.

1. An *expansion dialogue of length* n *for* A_{AF} given UAF is a nonempty sequence $X = X_0, \ldots, X_j, \ldots$ such that all of the following conditions hold:

   (a) $X_0 = AF$;

   (b) If $n \in \mathbb{N}$ then $X = X_0, \ldots, X_j$ and $0 \leq j \leq n$;

   (c) For every $i > 0$, X_i is an allowed expansion of X_{i-1} given UAF;

   (d) if i is odd then $A_{Xi} <_c^x A_{AF}$;

   (e) if i is even then $A_{Xi} \geq_c^x A_{AF}$.

   A finite expansion dialogue $X_0, \ldots, X_i$ of length n is *terminated* if either $n \in \mathbb{N}$ and $i = n$ or else there are no allowed expansions of X_i satisfying the above conditions.

2. An *expansion game* of length n about A_{AF} given UAF is a game with two players, a proponent and an opponent of A, who take turns after each move, where the proponent starts with AF. A move is legal if the resulting sequence of moves is an expansion dialogue of length n. A game is *terminated* iff it is terminated as a dialogue. A terminated game is *won* by the player who made the last move.

Let us consider some examples.

Example 3 Consider the AF and UAF in Figure 1. Clearly, in any game about B of any length, the proponent has a winning strategy AF, since UAF does not contain defeaters of B. However, in games about A things are more subtle. Assume first that an expansion is allowed iff it does not omit defeats from UAF. Then the opponent has a winning strategy in a game about A of length 1 but not in games of any length $n > 1$. The winning strategy for length 1 is AF, AF' where AF' expands AF with C and $A \leftarrow C$. However, in a game of any length $n > 1$ the proponent can reply by

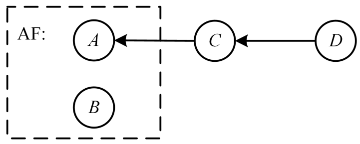

Figure 1: Illustrating the expansion game

extending AF' to AF'' with D and $C \leftarrow D$, after which the dialogue terminates with a win by the proponent.

Assume next that all expansions are allowed, so omitting defeats from UAF is allowed. Then the opponent has a winning strategy for all n, namely, AF, AF' where AF' expands AF with C, D and $A \leftarrow C$, so AF' omits $C \leftarrow D$ from UAF. The proponent cannot expand AF' since there are no new arguments to be introduced.

At first sight, the second variant of this example would seem to imply that it is easy to trivialise the game but this is not true. It all depends on the context provided by the UAF and on what the definition of allowed expansions implies about omitting defeats. Depending on the context and the nature of the application, omitting defeats may always, never or sometimes make sense. For example, in Section 5.2 we will in the context of *ASPIC*$^+$ give a sensible definition in which omitting asymmetric defeats is never allowed but omitting symmetric defeats may be allowed. For the same reasons it is in general undesirable to require that the expansions in an expansion game are minimal in their sets of arguments, since there is no a priori reason why expansions such as in the second variant of Example 3 should never be constructed.

The following relation can be proven between the existence of winning strategies in expansion games of length ∞ and the contextual status of an argument in an AF compared to in UAF.

Proposition 4 For any AF in a given UAF, if all expansions are allowed if and only if they do not omit defeats from UAF, then the following holds for all $A \in AF$.

1. If $A_{UAF} <^x_c A_{AF}$ then the opponent has a winning strategy in a game of length ∞ about A.

2. If the opponent has a winning strategy in a game of length ∞ about A, then $A_{UAF} \not\geq^x_c A_{AF}$.

PROOF. For (1), if $A_{UAF} <^x_c A_{AF} AF \neq UAF$, then AF, UAF is a winning strategy for the opponent in a games of length ∞ since there are no expansions of UAF.

For (2) assume that the opponent has a winning strategy S in a game of length ∞ about A. Suppose for contradiction that $A_{UAF} \geq^x_c A_{AF}$. Then the proponent can continue any dialogue X in S with UAF since $A_{X_i} <^x_c A_{AF}$ for any final move X_i in such a dialogue X. But then S is not a winning strategy for the opponent. Contradiction. QED

The condition that AF does not omit defeats from UAF is essential for this proposition. A simple counterexample without this condition is the following example.

Example 5 Consider a UAF consisting of A and B where $A \leftrightarrow B$ and an AF in UAF with the same arguments but with only $A \leftarrow B$. Then $A_{UAF} <^x_c A_{AF}$ but no expansion of AF exists, so so the opponent has no winning strategy in a game of length 1 about A.

It should be noted that the assumption that an expansion is allowed iff it does not omit defeats from UAF excludes many realistic applications of the model. In Section 5 we will see that instantiating the abstract model with *ASPIC*$^+$ imposes structural constraints and allows for sensible omissions of defeats. Moreover, there can be procedural constraints on expansions. For example, a judge could rule particular evidence or arguments inadmissible.

3.3 Attack Targets and Relevant Sets

To define defeat points, we also need notions of attack targets and relevant sets. First it is important to note that a defeat point must be defined as (at least) consisting of a *set* of arguments, as Example 6 shows.

Example 6 Consider Figure 2.

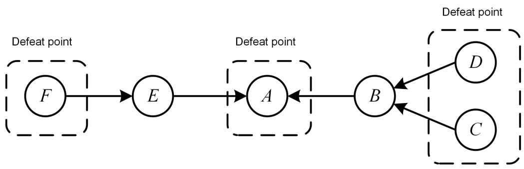

Figure 2: Multiple defeat points

Defeating just C or just D is not enough to lower the status of A, so $\{C, D\}$ should also be, or at least be part of, a defeat point. Note, furthermore, that defeating F also lowers the status of A, so $\{F\}$ should also be (part of) a defeat point of A, so an argument can have multiple defeat points.

This example suggests that defeat points must be defined as a set of arguments, but in fact the definition must be more refined. Imagine an AF with two defeatable but undefeated arguments A and B such that for both of them expansions exist that lower their status. Then on the account given so far they should both have one defeat point, namely, $\{A\}$, respectively, $\{B\}$. However, if A has just one premise on which it can be defeated while B has two, or A uses one defeasible rule while B uses two, and their status can be lowered by attacks on all these points, then A should still be dialectically stronger than B.

Example 7 A simple formal example in *ASPIC*$^+$ is the following two arguments, in which both p and r are ordinary premises:

A: $p \Rightarrow q$
B: $p, r \Rightarrow s$

Argument A has three points at which it can be attacked, namely, the ordinary premise p, the conclusion q and the application of the defeasible rule $p \Rightarrow q$, while B has four such 'attack targets', namely, the ordinary premises p and r, the conclusion s and the application of the defeasible rule $p, r \Rightarrow s$.

Can we leave the handling of this difference to structured instantiations of our account? No, since we want that observations made at the abstract level are inherited by instantiations, otherwise we run the risk of implicitly making assumptions that do not hold for all instantiations, such as that all arguments have the same number of 'weak spots'. Accordingly, we assume that each argument Y in a *UAF* comes with a finite set of attack targets, and we assume that each argument X defeating Y defeats Y on at least one of Y's attack targets. This will allow us to define defeat points as sets of argument-attack-target pairs.

Definition 17 [**Attack target function and attack targets**] Given a set T, an *attack target function* for a $UAF = (\mathcal{A}^u, \mathcal{D}^u)$ is a function $t : \mathcal{A}^u \longrightarrow 2^T$ that assigns to each argument in $\mathcal{A}^u$ a set of *attack targets*. Each UAF is assumed to have a unique attack target function. Moreover for each defeat relation (X, Y) from $\mathcal{D}^u$ a unique nonempty set $S \subseteq t(Y)$ is assumed consisting of the points *on which* X defeats Y.

For now the nature of the set T and the function t will be left implicit; we hope that Example 7 sufficiently illustrates the underlying ideas for the time being. In Section 5.3 we will give example instantiations for *ASPIC*$^+$.

Given a set S of arguments, we write S^t for the set of all pairs (A, t) such that $A \in S$ and $t \in t(A)$.

A further ingredient needed for defining defeat points is a notion of relevance of a set of defenders to the status of the defended argument. This notion is needed since it can happen that defeating a defender of A does not change the contextual status of A, as Example 8 shows.

Example 8 In the AF in Figure 3, C and G are defenders of A but defeating either of them does not lower the status of A; this only happens if either A or D is defeated, so, assuming that all arguments have undefeated defeaters in *UAF*, (which is left implicit) the only sets relevant to A are $\{A\}$ and $\{D\}$.

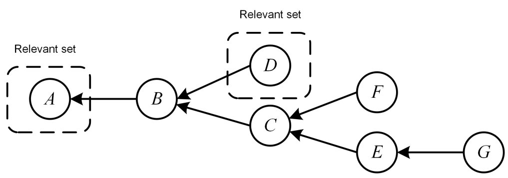

Figure 3: (Ir)relevant sets

Accordingly, the following definition (which adapts the dialogical notion of relevance proposed in [58] to AFs) says that a set $S \subseteq A$ is relevant to A in AF iff AF can be expanded with undefeated defeaters of all members of S such that A's contextual strength is lowered.

Definition 18 For any $AF = (\mathcal{A}, \mathcal{D})$ with $A \in \mathcal{A}$, a set $S \subseteq \mathcal{A}$ is *relevant to* A in AF iff S is a minimal set such that $A_{AF'} <_c^x A_{AF}$ for some $AF' = (\mathcal{A} \cup S', \mathcal{D}')$ such that:

- $S' \cap \mathcal{A} = \emptyset$; and
- $\mathcal{D}'$ consists of nothing else but defeat relations (B, B') for all $B \in S$ and some $B' \in S'$.

Note that this notion does not need to be defined relative to a *UAF*, in particular, S' does not need to be from *UAF*, since all the notion of relevance needs to capture is the sets of arguments to which an expansion that is meant to lower A's status should be targeted. For this reason a set S' as required by the definition always exists. Whether an expansion that lowers A's status is possible and, if so, whether it is allowed, depends on the content of a *UAF* and the definition of allowed expansions as used in Definition 19 below.

3.4 Defeat points

We can now formalise when an allowed expansion is successful in lowering the status of the focus argument. This will be done in terms of the existence of a winning strategy for the opponent in an expansion game. More precisely, a defeat point of an argument A is formally defined as a subset-minimal set S of argument-attack-target pairs that is relevant to A and for which an allowed expansion exists that defeats all attack targets in S and lowers A's contextual strength, and is the first opponent move in a winning strategy (in the usual game-theoretical sense) for the opponent in an expansion game. The definition is parametrised by a variable n for the maximum length of a game.

Definition 19 [**Defeat points**] Let $n \in \mathbb{N} \cup \{\infty\}$. Given an abstract argumentation framework $AF = (\mathcal{A}, \mathcal{D})$ in UAF, an *n-attack point* of an argument $A \in \mathcal{A}$ is any minimal set $S \subseteq \mathcal{A}^t$ relevant to A such that the opponent has a winning strategy in an expansion game of length n about A_{AF} given UAF, where the opponent's first move is $X_1 = (\mathcal{A}_1, \mathcal{C}_1)$ and is such that

- for all $(B, t) \in S$ there exists an argument $C \in \mathcal{A}_1$ such that C defeats B on t according to $\mathcal{D}_1$.

The set of n-defeat points of A given AF is denoted by $dp^n_{AF}(A)$.

By definition of the expansion game A cannot have n-defeat points for $n = 0$. Below we leave this implicit. Moreover, if S is a winning strategy for the opponent, then for the final move X_i of any terminated dialogue in S it holds that $A_{Xi} <_c A_{AF}$. Then it follows that the definition of 'attack points' in [62] is the special case of Definition 19 with $n = 1$, so our present approach formally generalises that of [62]. As for notation, when an argument A has a single attack target t, we will often say that A has n-defeat point $\{A\}$ instead of saying that it has n-defeat point $\{(A, t)\}$. Also, if there is no danger for confusion, we will speak of defeat points, leaving n implicit.

That $S^t \subseteq \mathcal{A}^t$ has to be minimal is for two reasons: to prevent that too many defeat points have to be computed, and to avoid trivialisation of our approach if *UAF* is infinite. As regards the first reason, consider an AF with a large set of arguments with undefeated defeaters in *UAF*, let $n = 1$ and let A be an argument in the AF with just one minimal 1-defeat point $\{A\}$. Then without the requirement of minimality, any subset of S^t containing A would be a 1-defeat point of A. As regards trivialisation in case of infinite AFs, without the requirement of minimality it may be that two arguments with different finite numbers of minimal n-defeat points both have an infinite number of non-minimal n-defeat points, which would render meaningful comparison impossible. The condition that the set of arguments of a defeat point of A is relevant to A is to exclude examples like the following one.

Example 9 Consider the *UAF* and AF in Figure 4, where A and B are assumed to have a single attack target. Without the relevance condition and if expanding AF with C is only allowed if both defeat relations are included in the expansion, then $\{B\}$ would, for any n, be an n-defeat point of A, which is undesirable.

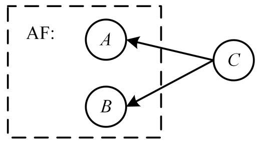

Figure 4: Why the set of arguments of defeat points should be relevant

Example 10 To further illustrate Definition 19, assume in Examples 6 and 8 that all arguments have a single attack target and that they all have undefeated defeaters in *UAF* outside AF. Then in Figure 2 the n-defeat points of A are, for any n, $\{A\}$, $\{F\}$ and $\{C, D\}$. This holds for all n since all expansions that satisfy Definition 19 are such that there is no reply to them in the expansion game. In Figure 3 the n-defeat points of A are $\{A\}$ and $\{D\}$, also for any n.

It is not required that all arguments in $\mathcal{A}'$ defeat some argument in S, since including a defeater 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 19 allows for 'side effects' in that the new defeaters may also defeat arguments outside S, or in that arguments in $\mathcal{A}$ but outside S may defeat them. For example, an argument defeating another argument on its premise will usually also defeat all other arguments using that premise (as in Example 2). Such side effects may be induced by an underlying structured account of argumentation. These points will be illustrated in more detail in Section 5 below.

The following proposition can be proven about the relation between being a 1-defeat point and being an n-defeat point for arbitrary n.

Proposition 11 For any AF in a given UAF and all $A \in AF$ it holds for all n that if S is an n-defeat point of A_{AF} then S is a 1-defeat point of A_{AF}.

PROOF. For any expansion dialogue in any winning strategy for the opponent in an expansion game of length n about A_{AF}, the opponent's first move satisfies Definition 19 by Definition 16 of an expansion dialogue. QED

However, the converse does not hold. Example 3 is a counterexample.

3.5 Dialectical Argument Strength: Ranking-based Semantics

We can now give our definition of dialectical argument strength, by combining the notion of contextual strength with the number of defeat 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. 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 defeat points. Moreover, we parametrise the definition with variables for the adopted contextual semantics and for the maximum length of expansion games.

Definition 20 **[Dialectical strength]** Let $AF = (\mathcal{A}, \mathcal{D})$ and $AF' = (\mathcal{A}', \mathcal{D}')$ be two abstract argumentation frameworks in a given *UAF* with semantics x and let $\geq_c^x$ be a contextual argument ordering for the set of all AFs in *UAF*. Let $n \in \mathbb{N} \cup \{\infty\}$. For any $A \in \mathcal{A}$ and $B \in \mathcal{A}'$ we say that $A_{AF} \geq_d^{x,n} B_{AF'}$ iff

1. $A_{AF} \geq_c^x B_{AF'}$; and

2. if $B_{AF'} \geq_c^x A_{AF}$ then $|dp^n_{AF}(A)| \leq |dp^n_{AF'}(B)|$.

The same notational conventions hold for $\geq_d^{x,n}$ as for $\geq_c^x$.

The following examples illustrate this definition.

Example 12 Consider the AFs in Figure 5 and the example $\geq_c^g$ for grounded semantics. Suppose a proponent of a claim φ has constructed an argument A with conclusion φ and now finds itself in a dialogue state with $AF_0 = A \leftarrow B$. The proponent considers two options to defend claim φ in the dialogue: defending A against B with a defeat C of B (resulting in AF_2), or constructing a new argument A' for conclusion φ (resulting in AF_1). Which move results in the dialectically stronger position? This amounts to comparing A_{AF2} and A'_{AF1}. Note that A and A' are both *in* so contextually they are

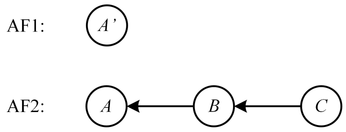

Figure 5: Comparing arguments in different AFs

equally strong, so what is decisive is how many defeat points they have. Many current gradual accounts regard A'_{AF1} as stronger than A_{AF2} based on the intuition that having no defeaters is better than having defeaters (the principle of Void Precedence discussed in Section 4.3 below). In our approach, this depends on several things. Suppose first that all of A, A', B and C have undefeated defeaters in *UAF* (not shown in the figure), that A, A' and C each have one attack target, respectively, t, t' and t'', that both A and A' have other undefeated defeaters in *UAF* besides B, and that all expansions are allowed. Finally, assume for simplicity that $n = 1$ (but the analysis holds for all n). Then A'_{AF1} has just one 1-defeat point, namely, $\{(A', t')\}$, while A_{AF2} has two 1-defeat points, namely, $\{(A, t)\}$ and $\{(C, t'')\}$. So $A'_{AF1} >_d^{g,1} A_{AF2}$, so in this case having no defeaters is strictly better.

However, assume now that both A and A' have no other defeaters in *UAF* besides B, or that they do have other defeaters in *UAF* but that no expansion with these other defeaters is allowed, perhaps for efficiency reasons. In both cases A'_{AF1} still has the single 1-defeat point $\{(A', t')\}$ but A_{AF2} now also has just one 1-defeat point, namely, $\{(C, t'')\}$. So $A_{AF2} \approx_d^{g,1} A'_{AF1}$, so here having no defeaters is not strictly better than having defeaters.

Finally, suppose that only A' has additional defeaters in *UAF* while A's only defeater in *UAF* is B. Then A_{AF2} has no 1-defeat points while A'_{AF1} has one 1-defeat point, namely, $\{(A', t')\}$. So $A_{AF2} >_d^{g,1} A'_{AF1}$, so we have a case where an argument that has defeaters is strictly better than an argument that has no defeaters. In conclusion, whether having no defeaters is better than having defeaters depends on the nature of the arguments and their relations and on the context in which they are evaluated.

Example 3 (Cont) Consider again the example $\geq_c^g$ for grounded semantics and assume first that expansions are allowed if and only if they do not omit defeat relations from UAF. Then since the opponent has a winning strategy in a game of length 1 about A, clearly $\{A\}$ is a 1-defeat point of A. Since B has no defeaters in UAF, B has no

n-defeat points for any n. Then $A_{AF} <_d^{g,1} B_{AF}$. However, for $n > 2$ and $n = \infty$ instead the proponent has a winning strategy, namely, by further extending AF' with D and $C \leftarrow D$, so A has no n-defeat points, so for $n > 2$ and $n = \infty$ we have that $A_{AF} \approx_d^{g,n} B_{AF}$.

Assume next that all expansions are allowed. Then for $n = 1$ the outcome is the same but for $n > 2$ and $n = \infty$ it is different. As explained above, the opponent then has a winning strategy in a game of length n about A by expanding AF with C and D and $A \leftarrow C$ but not $C \leftarrow D$. So now we have for all $n > 0$ that $\{A\}$ is an n-defeat point of A so for all $n > n$ it holds that $A_{AF} <_d^{g,n} B_{AF}$.

3.6 General Properties of Dialectical Argument Strength

In this section we investigate some general properties of Definition 20. In all four propositions that we prove we leave the superscripts that denote the semantics and game length used in defining contextual strength implicit, since the propositions hold for any semantics and game length.

First, $\geq_d$ can be shown to be a preorder on the set of all argument-AF pairs.

Proposition 13 For all arguments A, B, C and argumentation frameworks AF, AF' and AF'' in a given *UAF* it holds that:

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

PROOF. For (1) we have that $A_{AF} \approx_c A_{AF}$ since $\geq_c$ is assumed to be a preorder. Then also $A_{AF} \approx_d A_{AF}$ since $|dp_{AF}(A)| = |dp_{AF}(A)|$.

For (2), note first that if $A_{AF} \geq_c B_{AF'}$, then either $A_{AF} >_c B_{AF'}$ or $A_{AF} \approx_c B_{AF'}$. Now assume that $A_{AF} \geq_d B_{AF'}$ and $B_{AF'} \geq_d C_{AF''}$. Then the following cases must be considered.

If $A_{AF} >_c B_{AF'}$, then if $B_{AF'} >_c C_{AF''}$ then $A_{AF} >_c C_{AF''}$ by transitivity of $\geq_c$, so $A_{AF} >_d C'_{AF'}$. If instead $B_{AF'} \approx_c C_{AF''}$, then $A_{AF} \geq_c C_{AF''}$ by transitivity of $\geq_c$. If also $C_{AF''} \geq_c A_{AF}$ then $B_{AF'} \geq_c A_{AF}$ by transitivity of $\geq_c$. Contradiction. So $A_{AF} >_c C_{AF''}$ but then $A_{AF} >_d C_{AF''}$.

If $A_{AF} \approx_c^x B_{AF'}$, then if $B_{AF'} >_c C_{AF''}$ then $A_{AF} \geq_c C_{AF''}$ by transitivity of $\geq_c$. If also $C_{AF''} \geq_c A_{AF''}$ then $C_{AF''} \geq_c B_{AF'}$ by transitivity of $\geq_c$, which contradicts that $B_{AF'} >_c C_{AF''}$. So $A_{AF} >_c C_{AF''}$ but then $A_{AF} >_d C_{AF''}$.

Finally, if $A_{AF} \approx_c B_{AF'}$ and $B_{AF'} \approx_c C_{AF''}$ then $A_{AF} \approx_c C_{AF''}$ since $\geq_c$ is a preorder. Then $|dp_{AF}(A)| \geq |dp_{AF'}(B)|$ and $|dp_{AF'}(B)| \geq |dp_{AF''}(C)|$. But then $|dp_{AF}(A)| \geq |dp_{AF''}(C)|$ so $A_{AF} \geq_d^x C_{AF''}$. QED

Next, strict contextual preference implies strict dialectical preference.

Proposition 14 If $A >_c B$ then $A >_d B$.

PROOF. First, $A_{AF} \geq_d B_{AF'}$ holds since $A_{AF} >_c B_{AF'}$ implies $A_{AF} \geq_c B_{AF'}$, so condition (1) of Definition 20 is satisfied, and $A_{AF} >_c B_{AF'}$ implies $B_{AF'} \not\geq_c A_{AF}$, so condition (2) of Definition 20 is satisfied. Next, $B_{AF'} \geq_d A_{AF}$ does not hold since $A_{AF} >_c B_{AF'}$ implies $B_{AF'} \not\geq_c A_{AF}$, so condition (1) of Definition 20 is not satisfied. QED

Next, if no expansions are possible, then dialectical preference reduces to contextual preference.

Proposition 15 If no AF in *UAF* has an expansion in *UAF*, then $A \geq_d B$ if and only if $A \geq_c B$.

PROOF. Immediate from Definition 20. QED

Finally, the ordering of dialectical strength is total if and only if the embedded ordering of contextual strength is total.

Proposition 16 $\geq_d$ is total if and only if $\geq_c$ is total.

PROOF. The only-if part follows since if $A \geq_d B$ then $A \approx_c B$ or $A >_c B$. The if-part follows from the fact that $\leq_c$ has primacy in Definition 20 that for $\approx_c$ cases it is refined into $\leq_d$ by an ordering on the cardinality of sets, which is total. QED

We next investigate relations between the special case of [62] and the present general case.

First, Example 3 (Cont) illustrated that the following does not hold in general: for all n: if $A_{AF} \geq_d^{x,n} B_{AF'}$ then $A_{AF} \geq_d^{x,1} B_{AF'}$. In the case in which expansions are allowed iff they do not omit defeat relations from UAF we had $A_{AF} \geq_d^{g,2} B_{AF}$ but $A_{AF} \not\geq_d^{g,1} B_{AF}$.

Next, the following modification of the example shows that it also does not hold in general for all $n > 0$ that if $A_{AF} \geq_d^{x,1} B_{AF'}$ then $A_{AF} \geq_d^{x,n} B_{AF'}$.

Example 17 Extend Example 3 as in Figure 6 and assume that expansions are allowed iff they do not omit defeat relations from UAF. Then A has a single 1-defeat point $\{A\}$ and B has a single 1-defeat point $\{B\}$ which is also a 2-defeat point of B. However, $\{A\}$ has no 2-defeat point. Suppose the opponent expands AF to AF' by including C and $A \leftarrow C$. Then the proponent can win by further extending AF' with D and $C \leftarrow$

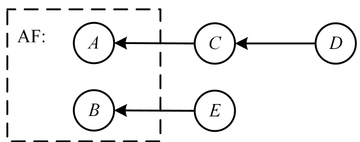

Figure 6: Dialectical strength with expansion games of varying length

D. So $A_{AF} \approx_d^{g,1} B_{AF}$ but $A_{AF} >_d^{g,2} B_{AF}$, hence $B_{AF} \geq_d^{g,1} A_{AF}$ but $B_{AF} \not\geq_d^{g,n} A_{AF}$.

Finally, it might be thought that if all expansions are allowed, then dialectical strength would equate contextual strength in the limit, for instance, if $A_{UAF} <_c B_{UAF}$ then $A_{AF} <_d^{x,\infty} B_{AF'}$. However, this does not hold. The following example is a simple counterexample.

Example 18 Consider the AF and UAF in Figure 7, where both A and B have a single attack target, let $x = g$ and assume that all expansions are allowed. Then we have that $A_{UAF} <_c B_{UAF}$ but both for A and for B the opponent has a winning strategy in a game of any length n by expanding AF with, respectively, C and $A \leftarrow C$ (in a game about A) and D and either $B \leftarrow D$ or $B \leftrightarrow D$ (in a game about B), after which the game terminates. So $A_{AF} \approx_d^{g,\infty} B_{AF}$.

However, the following relation can be proven.

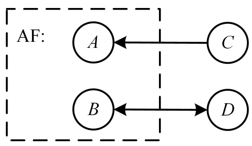

Figure 7: Dialectical strength versus contextual strength (1)

Proposition 19 For all AF and AF' in a given UAF, if all expansions are allowed iff they do not omit defeats from UAF, then the following holds for all $A \in AF$ and $B \in AF'$: if $A_{UAF} <_c A_{AF}$ while $B_{UAF} \geq_c B_{AF'}$, then for all $n > 1$: $A_{AF} <_d^{x,n} B_{AF'}$.

PROOF. Note first that if $A_{UAF} <_c A_{AF}$ then $AF \neq UAF$, so AF, UAF is a winning strategy for the opponent in all expansion games about A of any length. But then for all such n it holds that A_{AF} has at least one n-defeat point. Moreover, for all games AF', X_1 about B we have that $X_1 \neq UAF$ since $B_{X1} <_c B_{AF}$ while $B_{UAF} \not<_c B_{AF'}$. But then for all such games UAF is a continuation that terminates the dialogue with a win for the proponent, so for no $n > 1$ does it hold that $B_{AF'}$ has an n-defeat point. But then for all $n > 1$: $A_{AF} <_d^{x,n} B_{AF'}$. QED

The following is a counterexample to Proposition 19 for $n = 1$.

Example 20 Consider the AF and UAF in Figure 8, where both A and B have a single attack target, let $x = g$ and assume that all expansions are allowed iff they do not omit defeats from UAF. Then all other conditions of Proposition 19 hold but

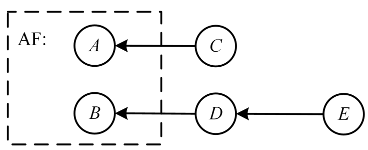

Figure 8: Dialectical strength versus contextual strength (2)

$A_{AF} \approx_d^{1,n} B_{AF}$, since both for A and for B the opponent has a winning strategy in a game of length 1 by expanding AF with, respectively, C and $A \leftarrow C$ (in the game about A) and D and $B \leftarrow D$ (in the game about B), after which the game terminates.

4 Dialectical Argument Strength with Particular Contextual Argument Orderings

In this section we explore the use of various particular contextual argument orderings in Definition 20 of dialectical argument strength.

4.1 Contextual Strength with Dung's Semantics for Abstract Argumentation Frameworks

We first consider a generalisation of the above contextual ordering for grounded semantics to any of the four semantics defined in Dung's seminal (1995) paper. Several such generalisations are possible and the following one, stated in terms of labellings, seems a reasonable one and arguably improves two existing proposals in some respects. Let us first, given an AF and $A \in AF$ and a semantics $x \in$ {grounded, preferred, stable, complete}, define the following so-called *labelling classes* of arguments:

- $A \in \forall$ iff A is *in* in all x-labellings of AF and there exists such a labelling;
- $A \in \exists\exists\neg\exists$ iff A is *in* in some x-labellings, *und* in some x-labellings but *out* in no x-labellings of AF;
- $A \in \exists?\exists$ iff A is *in* in some x-labellings and *out* in some x-labellings of AF;
- $A \in \forall^{=}$ iff A is *und* in all x-labellings of AF and there exists such a labelling;
- $A \in \neg\exists\exists\exists$ iff A is *in* in no x-labellings, *und* in some x-labellings and *out* in some x-labellings of AF;
- $A \in \forall^{\emptyset}$ iff AF has no x-labellings;
- $A \in \forall^{\neg}$ iff A is *out* in all x-labellings of AF and there exists such a labelling.

It is easy to see that every argument is in exactly one of these classes. Then $\geq_c^x$ is a *classic contextual ordering* iff $x \in$ {grounded, preferred, stable, complete} and $\geq_c^x$ is the transitive closure of the following statements (visualised in Figure 9):

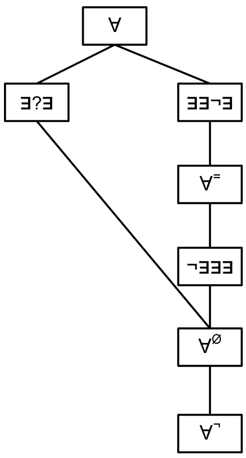

Figure 9: The classic contextual ordering

- If A_{AF} and $B_{AF'}$ are in the same class, then $A_{AF} \approx_c^x B_{AF'}$
- If $A_{AF} \in \forall$ and $B_{AF'} \notin \forall$ then $A_{AF} >_c^x B_{AF'}$
- If $A_{AF} \in \forall^{\neg}$ and $B_{AF'} \notin \forall^{\neg}$ then $B_{AF'} >_c^x A_{AF}$
- If $B_{AF'} \in \forall^{\emptyset}$ and $A_{AF} \notin \forall^{\emptyset} \cup \forall^{\neg}$ then $A_{AF} >_c^x B_{AF'}$
- If $A_{AF} \in \forall^{=}$ and $B_{AF'} \in \neg\exists\exists\exists$ then $A_{AF} >_c^x B_{AF'}$
- If $A_{AF} \in \exists\exists\neg\exists$ and $B_{AF'} \in \forall^{=}$ then $A_{AF} >_c^x B_{AF'}$

Note that the example contextual ordering for grounded semantics mentioned in Section 3.1 is a special case of this definition.

There are, to the best of our knowledge, two alternative proposals for ordering arguments in terms of labelling semantics. In [3] an ordering is proposed in which $\exists\exists\neg\exists$ and $\exists?\exists$ are one class. Let us denote this class with $\exists??$. Moreover, $\neg\exists\exists\exists$ and $\forall^{\neg}$ are also one class. Let us denote this class with $\neg\exists?\exists$. Finally, in [3] $\forall^{=}$ and cases with no labellings are regarded as one class. Let us denote this with $\forall^{=,\emptyset}$. The contextual ordering is then induced by $\forall > \exists?? > \forall^{=,\emptyset} > \neg\exists?\exists$ and is total. Note that the present ordering is not just a refinement of [3]'s ordering, since [3] has $\forall^{=} < \exists?\exists$ while in the present ordering the relation between these classes is undefined. The latter seems desirable since $\exists?\exists$ is better than $\forall^{=}$ in that the argument is *in* in at least one labelling, but worse in that the argument is *out* in at least one labelling. So arguments in these two classes arguably are of incomparable contextual strength.

In [75] an ordering is proposed for complete semantics, so without the class $\forall^{\emptyset}$. Apart from this, the only difference with our ordering is that while we have an undefined relation between $\exists\exists\neg\exists$ and $\exists?\exists$, in [75] the first class is better then the second. A similar line of reasoning as above motivates our design choice. Suppose we have two arguments A and B from the same AF which has three labellings l_1, l_2 and l_3 and suppose we have

	l_1	l_2	l_3
A	*in*	*in*	*out*
B	*und*	*in*	*und*

We see that A has a better label than B in l_1 but a worse label in l_3. So all in all they arguably are incomparable. The undefined relation between $\neg\exists\exists\exists$ and $\exists?\exists$ can be justified in the same way. Finally, the place of $\forall^{\emptyset}$ is motivated by the fact that if an AF has no labelling (which can only happen in stable semantics) then the only way to lower the status of an argument is to make it *out* in all labellings, which seems reasonable.

Having said so, it would be interesting to explore the effect of differences in the choice of $\geq_c^x$ for the classic semantics with respect to the results reported below.

The following result involves a so-called defeat property that an AFs in a *UAF* can have.

Definition 21 An AF in a given *UAF* satisfies the *defeat property* iff for all arguments A, B and C in $\mathcal{A}_{AF}$ and for all attack targets t that are shared by A and B it holds that C defeats A on t in AF iff C defeats B on t in AF.

Proposition 21 Let AF and all its allowed expansions satisfy the defeat property, let $A, B \in \mathcal{A}_{AF}$ and consider a classic contextual ordering $\geq_c^x$. Then if $t(A) \subseteq t(B)$ then for all n:

1. $A \not<_d^{x,n} B$.
2. If $\geq_c^{x,n}$ is total then $A \geq_d^{x,n} B$.

PROOF. Proof of (1). Suppose for contradiction that $A <_d^{x,n} B$ for some n. We first prove that (i) $A \approx_c^x B$. Suppose for contradiction first that $\geq_c^x$ is undefined for A and B. Then also $\geq_d^{x,n}$ is undefined for A and B. Contradiction.

Suppose next that $A <_c^x B$. If $A \in \forall^{\neg}$, then there exists a labelling for AF so $B \notin \forall^{\emptyset}$. Moreover, in any labelling there exists a defeater C of A that is *in*. But then by the defeat property any such C also defeats B so B is *out* in all labellings, so $A \approx_c^x B$. So $A \notin \forall^{\neg}$.

If $A \in \forall^{\emptyset}$, then there exists no labelling for AF so $B \in \forall^{\emptyset}$, so $A \approx_c^x B$. So $A \notin \forall^{\emptyset}$.

If $A \in \neg\exists\exists\exists$ then there exists a labelling of AF in which A is *out*. Then there exists a defeater C of A that is *in* in that labelling so B is *out* in that labelling. So $B \in \exists?\exists$. But then there exists a labelling of AF in which B is *in* so all defeaters of B

are *out* in that labelling. But among these are also all defeaters of A, so A is *in* in that labelling. But then $A \approx_c^x B$ so $A \notin \neg\exists\exists\exists$.

If $A \in \forall^=$ then in all labellings of AF there exists a defeater of A that is *und* and there exists no defeater of A that is *in*. Then in all such labellings there exists a defeater of B that is *in* or *und* so in all such labellings B is *out* or *und*. Then $A_{AF} \geq_d^{x,n} B_{AF}$ so $A \notin \forall^=$.

Finally, if $A \in \exists?\exists$ or $A \in \exists\exists\neg\exists$ then there exists a labelling of AF in which some defeater of A is not *out*. But this defeater also defeats B, so $B \notin \forall$ so B is in the same labelling class as A, so $A \approx_c^x B$. Contradiction.

This proves that $A \approx_c^x B$. It implies that A and B are in the same labelling class. We must now prove that all n-defeat points of A are also an n-defeat point of B.

We first prove (ii) that for all n, all winning strategies for the opponent in a game of length n about A are winning strategies for the opponent in a game of length n about B. Suppose for contradiction that there is a winning strategy S for the opponent about A that is not a winning strategy for the opponent about B. Then for some minimal dialogue $X_1, \ldots, X_i$ in S it holds that the proponent has a reply X_{i+1} to X_i in the game about B but not in the game about A.

Consider any such reply X_{i+1} of the proponent in the game about B. It is such that $B_{Xi+1} >_d^{x,i+1} B_{Xi}$ and $B_{Xi+1} \geq_d^{x,i+1} B_{AF}$. For any labelling of X_{i+1} it holds that if B is *in* in X_{i+1} then all defeaters of B are *out* in X_{i+1}. But since $A^- \subseteq B^-$ by the defeat property, also all defeaters of A are *out* in X_{i+1}. But then A is *in* in X_{i+1}. Next, if B is *und* in X_{i+1} then no defeater of B is *in* in X_{i+1}. But then no defeater of A is *in* in X_{i+1}, so A is *und* or *in* in X_{i+1}. So in all labellings of X_{i+1} the labelling value of A is at least as preferred as that of B. But then it is straightforward to prove with Figure 9 that A is in a labelling class that is at least as preferred as $B's$ labelling class, so $A \geq_d^{x,i+1} B$. But then X_{i+1} is also a valid reply for the proponent in the game about A. Contradiction.

Next we must prove (iii) that for every relevant set A targeted by X_1 in a winning strategy for the opponent about A, either S is a relevant set targeted by X_1 in the game about B, or some subset S' of S that is not a relevant set for A is a relevant set for B. Then it follows from (ii) and (iii) that B has at least the same number of n-defeat points as A. To prove this, note that defeating S also reduces the contextual status of B. This holds since for any labelling of X_1, If A is *out* in X_1 then it has a defeater that is *in* in X_1 but then B also has this defeater so B is also *out* in X_1. In the same way, if A is *und* in X_1 then both A and B a defeater that is not *out* in X_1 so B is not *in* in X_1. But then if S is not minimal in the sense of Definition 19, some proper subset S' of S is minimal in this sense so S' is relevant to B. So B has at least the same number of n-defeat points as A, so $A \not<_d^{x,n} B$.

The proof of (2) immediately follows from the proof of (1) since the case where $\geq_c^{x,n}$ is undefined for A and B does not arise. QED

The following example illustrate that at the end of the proof we could not simply assume that all sets relevant to A defeated by X_1 are also relevant to B.

Example 22 Consider the modification of Example 2 depicted in Figure 10 (where B, C and D are duplicated for visual clarity). Then $\{C, D\}$ is not relevant to A' since $\{C\}$ is relevant to A'. This outcome seems reasonable since there is no way to reduce A''s contextual status by defeating C and D without defeating C.

Proposition 21 is what one would expect from dialectical strength as degree of vulnerability of an argument, since it says that extending an argument with additional attack

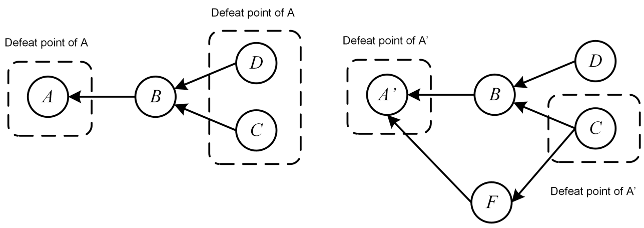

Figure 10: Non-minimal 'defeat points'

targets makes the new argument at best equally strong as the extended argument and possibly weaker. Its more general version where A and B can be from different AFs does not hold. A counterexample is displayed in Figure 11. Here $\{(A, t)\}$ is (for a

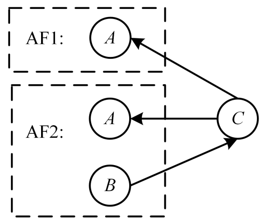

Figure 11: Counterexample to general version of Proposition 21

given t and all n) an n-defeat point 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.

4.2 Positive Results on Principles for Ranking-based Semantics for the Classic Contextual Ordering

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 [11], which is a comprehensive study of principles for ranking-based semantics. In this section we confine ourselves to the classic contextual ordering. 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.

Most principles discussed in [11] evaluate arguments in a single AF. Therefore we can only verify these principles for the special case that $AF = AF'$ in Definition 20. Furthermore, below we leave the superscripts of $\geq_c$ and $\geq_d$ implicit if there is no danger for confusion. In definitions the omission indicates that the notion is defined for any pair of semantics x and expansion dialogue length n while in propositions it indicates that the proposition holds for all such pairs. It turns out that of all principles discussed by [11], Definition 20 with classic contextual strength only satisfies the following principle, which say that an argument without defeat[6] branch is ranked higher than an argument only defeated by one undefeated argument.

Defeat vs Full Defense says for acyclic AFs that if argument $A \in AF$ has no defeat branches and argument $B \in AF$ is defeated by only one argument, which moreover, is undefeated in AF, then $A >_d B$.

Proposition 23 Definition 20 of dialectical argument strength with classic contextual strength satisfies Defeat vs Full Defense.

PROOF. This holds since any argument that has no defeat branch is *in* while any argument only defeated by one non-defeated argument is *out* in any labelling for any of the four classic semantics. Note that since AFs are assumed to be acyclic, in all these semantics at least one such labelling exists. QED

Counterexamples to the other principles will be given in Section 4.3. For three of the principles that do not hold in general, we have verified special cases in which they hold, while for one further principle we have identified a weaker version that holds in special cases.

Total says that all pairs of arguments can be compared: for any AF and all $A, B \in AF$: $A \geq_d B$ or $B \geq_d A$.

Proposition 16 stated a special case in which the Total principle holds, namely, when $\geq_c$ is total.

Quality Precedence says that the higher the rank of a defeater of A in an AF, the lower the rank of A. Formally:

> For all $A, B \in AF$, if there exists a $C \in B^-$ such that for all $D \in A^-$ it holds that $C >_d D$, then $A >_d B$.

Quality precedence holds in the following special case.

Proposition 24 Let $\geq_c$ be determined by grounded or stable semantics. Then for any AF and any $A, B, C \in AF$ such that C defeats B and $C >_c D$ for all defeaters D of A, it holds that $A >_d B$.

PROOF. Consider any AF and any $A, B, C \in AF$ such that C defeats B and for all $D \in A^-$ it holds that $C >_c D$. For grounded semantics the classic contextual ordering reduces to $\forall >_c \forall^= >_c \forall^\neg$. Suppose first that $C \in \forall$. Then $B \in \forall^\neg$. Moreover, no defeater D of A is in $\forall$, so $A \notin \forall^\neg$. So $A >_c B$ so $A >_d B$. Suppose next that $C \in \forall^=$. Then $B \in \forall^= \cup \forall^\neg$. Moreover, all defeaters D of A are in $\forall^\neg$, so $A \in \forall$ so $A >_c B$ so $A >_d B$.

For stable semantics the classic contextual ordering reduces to $\forall >_c \exists ? \exists >_c \forall^\neg$ given the assumption that $C >_c B$, which implies that there exists at least one stable labelling. Then the proof is similar as for grounded semantics. QED

[6] In [11] ‘attack’ is used where this paper uses ‘defeat’.

For a counterexample to the special case of Quality Precedence for preferred and complete semantics under the assumptions of Proposition 24, consider the AF in Figure 12. It has two preferred labellings, one with E *in* and E' *out* and one with E' *in* and E *out*. In the first labelling D is *out* and D' is *und* so A is *und*. In the second labelling D' is *out* and D is *und* so A is *und*. Moreover, in both these labellings B and C are *und*. So

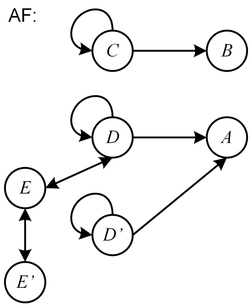

Figure 12: Counterexample to Quality Precedence for preferred and complete semantics

we have that all defeaters of A are inferior to C according to $\geq_c$ since $C \in \forall^=$ while all defeaters of A are in $\neg\exists\exists\exists$, while yet $A \approx_c B$. Then if both A and B have a single attack target and if UAF further only consists of undefeated defeaters of A and B, we have our counterexample.

Furthermore, we have proven a special case of **Distributed Defense Precedence**. This principle assumes definitions that the defense of an argument A is *simple* iff every direct defender of A defeats exactly one defeater of A, and *distributed* iff every defeater of A has at most one defeater. Then Distributed Defense Precedence says:

> If $|A^-| = |B^-|$ and $|A^+| = |B^+|$, and the defense of A is simple and distributed while the defense of B is simple but not distributed, then $A_{AF} >_d B_{AF}$.

Proposition 25 Definition 20 of dialectical argument strength with classic contextual strength satisfies Distributed Defense Precedence for any argument $A \notin \forall^\neg \cup \forall^\emptyset$.

PROOF. Consider an AF with $A, B \in AF$ such that $A \notin \forall^\neg \cup \forall^\emptyset$, and let x be any of the four classic semantics. Then there exists an x labelling in which A is not *out*, so all defeaters of A in AF have a defeater in AF. But then if the defense of both A and B is simple while the defense of A is distributed, then all defeaters of A have exactly one defeater, so if A and B have the same numbers of defeaters and defenders, the defense of B can only be not distributed if one of its defeaters has no defeaters (since another of its defeaters has at least two defeaters). But then B is *out* in all x labellings so $A >_c B$ so $A >_d B$. QED

Finally, the following weak version holds of Defense Precedence (see Section 4.3).

Weak Defense Precedence says that if A_{AF} and B_{AF} have the same number of defeaters in AF but A_{AF} has direct defenders in AF while B_{AF} has no direct defenders in AF, then B is not strictly stronger than A. Formally:

> If $|A^-| = |B^-|$ and $A^+ \neq \emptyset$ and $B^+ = \emptyset$ then $A_{AF} \not<_d B_{AF}$.

Proposition 26 Definition 20 with classic contextual strength satisfies Weak Defense Precedence for grounded, preferred and complete semantics but not for stable semantics.

PROOF. This holds since a defeated argument with no defenders is *out* in all grounded, preferred and complete labellings, while for all these semantics at least one labelling exists. So $B_{AF} \in \forall^{\neg}$, so $A_{AF} \not<_d B_{AF}$.

A counterexample for stable semantics is an AF with two arguments A and B where B defeats itself, and *UAF* containing a single further argument C that defeats A. Then both A and B are in $\forall^{\emptyset}$, so $A \approx_c B$. Moreover, A has one defeat point, namely, $\{A\}$, while B has no defeat points, since, even though $\{B\}$ is relevant to B and extending AF with C lowers the contextual strength of B to $\forall^{\neg}$, C does not defeat B. So $A <_d B$. QED

4.3 Negative Results on Principles for Ranking-based Semantics for the Classic Contextual Ordering

In this section we give counterexamples to principles proposed in the literature on ranking-based semantics, again based on [11] and for the case of classic contextual argument strength. Note that these counterexamples also hold for the general case of Section 3. All of the given counterexamples also hold for the special case where all considered expansions are allowed, and they hold for all four classic contextual semantics. Note that many counterexamples that at first sight would only seem to hold for length $n = 1$ of the expansion dialogue game, hold for all n since the expansions that they contain as the first and only move of the opponent have no further expansions that the proponent could move in reply.

Recall that most principles proposed in the literature are meant for evaluating arguments in a single AF and can therefore only be verified for the special case that $AF = AF'$ in Definition 20. Unless indicated otherwise, this special case is assumed below.

Abstraction says that different AFs of the same form should evaluate arguments having the same structural relations in the AFs equally. For a counterexample in all four classic semantics, consider AF_1 with just A and AF_2 with just B and both A and B having one attack target, where *UAF* further consists of C defeating B. Abstraction says that A and B are of the same rank but we have $A >^n_d B$ for all n.

Void Precedence says that a non-defeated argument is ranked strictly higher than any defeated argument in the same AF. Formally:

If $A^- = \emptyset$ and $B^- \neq \emptyset$, then $A_{AF} >_d B_{AF}$.

A counterexample can be constructed from Example 12 by merging AF_1 and AF_2, letting $A' = A_{AF}$ and $A = B_{AF}$. It holds for all n and for all four classic semantics.

Independence says that the ranking between two arguments in the same AF should be independent of arguments that are connected to neither. More formally [11] this is expressed in terms of the notion $cc(AF)$ of the *connected components* af an AF, which is the set of largest subgraphs of AF, where two arguments are in the same component of AF if and only if there is some path (ignoring the directions) between them. Then independence means:

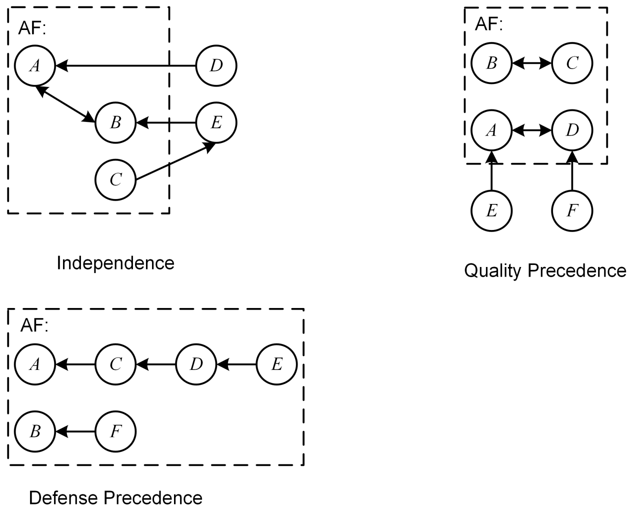

Figure 13: Counterexamples for ranking-based with classic contextual ordering (1)

> For all $AF' \in cc(AF)$ and all $A, B \in \mathcal{A}_{AF'}$: if $A_{AF'} \geq_d B_{AF'}$ then $A_{AF} \geq_d B_{AF}$.

This does not hold. A counterexample for all four classic semantics is in the top-left part of Figure 13. Here $cc(AF) = \{AF', AF''\}$ where $AF' = A \leftrightarrow B$ while $AF'' = C$ and all arguments have a single attack target. Then in AF' we have for all n that $\{A\}$ is an n-defeat point of A and $\{B\}$ is an n-defeat point of B so $A_{AF'} \approx_d B_{AF'}$. However, in AF, while $\{A\}$ still is an n-defeat point of A, argument B has no n-defeat points, since C protects B against E, so there is no expansion that lowers B's status. So $B_{AF} >_d^n A_{AF}$ for all n.

Self-Contradiction says that a self-defeating argument is ranked strictly lower than any non-self-defeating argument in the same AF. A counterexample for all four semantics is an AF with A defeating itself and C defeating B. Then in grounded, preferred and complete semantics $A >_c B$ so $A >_d B$ while in stable semantics $A \approx_c B$. Then if, for any n, A and B have the same number of n-defeat points, then $A \approx_d^n B$. Rather than considering changes of our definition to satisfy Self-contradiction, we leave the treatment of self-defeating arguments to structured accounts of contextual argument strength, since there may be defensible semantic reasons why self-defeating arguments should not always have the lowest contextual strength.

Cardinality Precedence says if A has a strictly greater number of direct defeaters than B in the same AF, then A is strictly weaker than B. Formally:

> If $|A^-| > |B^-|$ then $B_{AF} >_d A_{AF}$.

This holds for no n. The counterexamples in Example 12 to Void Precedence are also counterexamples to Cardinality Precedence.

Quality Precedence, for which Proposition 24 stated a special case in which it holds, does not hold in general. A counterexample for all four classic semantics and all n is in the top-right part of Figure 13. For all n it holds that $C >_d^n D$ since D has $\{D\}$ as n-defeat point while C has no n-defeat points. However, it also holds that $B >_d^n A$ since A has $\{A\}$ as n-defeat point while B has no n-defeat points. So $A \not>_d^n B$.

Defense Precedence says that if two arguments A and B in the same AF have the same number of defeaters but A has direct defenders while B has no defender, then A is strictly stronger than B. Formally:

If $|A^-| = |B^-|$ and $|A^+| \neq \emptyset$ and $|B^+| = \emptyset$ then $A >_d B$.

This does not hold for any n, not even when dialectical strength reduces to contextual strength. A counterexample for all four classic semantics and all n is in the bottom part of Figure 13. Assume that all arguments have a single attack target. Then both A and B are *out* so $A \approx_d^n B$ for all n.

Counter-Transitivity says that if the direct defeaters of B are at least as numerous and strong as those of A in the same AF, then A is at least as strong as B. Formally, let $\geq_S$ be a ranking on a set of arguments $\mathcal{A}$. For any $S_1, S_2 \subseteq \mathcal{A}$, $S_1 \geq_S S_2$ is a *group comparison* iff there exists an injective mapping f from S_2 to S_1 such that for all $A \in S_2$ it holds that $f(A) \geq_d A$. And $S_1 >_S S_2$ is a *strict group comparison* iff $S_1 >_S S_2$ and ($|S_2| < |S_1|$ or there exists an $A \in S_2$ such that $f(A) \geq_d A$). Then Counter-Transitivity says:

$$B^- \geq_S A^- \Rightarrow A \geq_d B$$

The counterexample against Defense Precedence in Figure 13 also holds against Counter-Transitivity, since Counter-Transitivity implies Defense Precedence [2].

Strict Counter-Transitivity says that if the defeaters of B are either more numerous or stronger than those of A in the same AF, then A is strictly preferred over B. Formally:

$$B^- >_S A^- \Rightarrow A >_d B$$

The counterexample against Defense Precedence also holds against Strict Counter-Transitivity, since Strict Counter-Transitivity implies Defense Precedence.

Distributed Defense Precedence, for which Proposition 25 stated a special case in which it holds, does not hold in general. A counterexample is displayed in Figure 14. The defense of A is simple and distributed while the defense of A' is simple but not

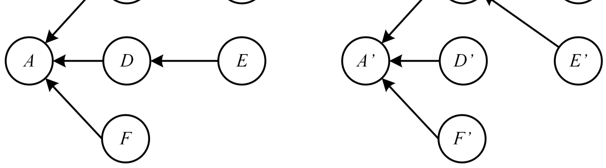

Figure 14: Counterexample to Distributed Defense Precedence

distributed. Moreover, A and A' have the same number of defeaters and the same number of defenders. Yet both are *out* in all labellings for all four classic semantics, so for all n we have $A \approx_d^n B$.

Non-Defeated Equivalence says that all non-defeated arguments are of equal strength:

If $A^- = B^- = \emptyset$ then $A_{AF} \approx B_{AF}$.

The above counterexample to Abstraction is also a counterexample to Non-defeated equivalence.

The following five principles can be verified in general since they compare different AFs. The first four of them consider expansions that consist of adding a defense or

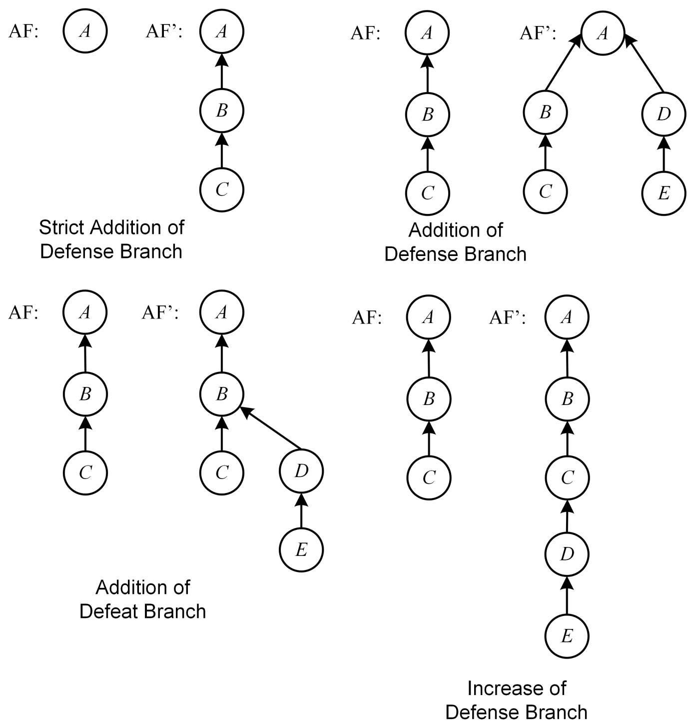

Figure 15: Counterexamples to Addition and Increase Principles

defeat branch to an isomorphic copy AF' of an AF. Note that a special case is when $AF' = AF$ and $A' = A$, so when the isomorphism is from AF to itself and it relates all arguments and defeat relations to themselves. Since for all four principles there are counterexamples for this special case, we will for simplicity formulate the principles for this special case. They assume the following definition from [11], adapted to our notation.

Definition 22 Let $AF = (\mathcal{A}, \mathcal{D})$ and $A \in \mathcal{A}$. The *defense branch added to* A is $P_+(A) = (\mathcal{A}', \mathcal{D}')$ with $\mathcal{A}' = \{x_0, \ldots, x_n\}$, $n \in 2\mathbb{N}$, $x_0 = A$, $\mathcal{A} \cap \mathcal{A}' = \{A\}$, and $\mathcal{D}' = \mathcal{D} \cup \{(x_i, x_{i-1}) \mid i \leq n\}$. The *defeat branch added to* A, denoted $P_-(A)$, is defined similarly except that the sequence is of odd length, i.e. $n \in 2\mathbb{N} + 1$.

Strict Addition of Defense Branch says that adding a defense branch to any argument improves its ranking.

If $AF' = AF \cup P_+(A)$ then $A_{AF'} >_d A_{AF}$

Strict Addition of Defense Branch does not hold. A counterexample is displayed in the top-left part of Figure 15. For both AF and AF' we have that $A \in \forall$ for all four classical semantics. If, moreover, all three arguments have one attack target and have undefeated defeaters in allowed expansions, then A in AF has for all n just one n-defeat point, viz. $\{A\}$, while A in AF' has an additional n-defeat point $\{C\}$. So $A_{AF} >^n_d A_{AF'}$ for all n.

Addition of Defense Branch says that adding a defense branch to any attacked argument improves its ranking:

If $AF' = AF \cup P_+(A)$ and A has defeaters in AF, then $A_{AF'} >_d A_{AF}$

This principle also does not hold. A counterexample is displayed in the top-right part of Figure 15. If all arguments except C have defeaters in *UAF* and if all these defeaters are undefeated, then (assuming all arguments have a single attack target) A in AF has, for all n, just one n-defeat point, viz. $\{A\}$, while A in AF' has an additional n-defeat point $\{E\}$. So $A_{AF} >_d A_{AF'}$.

Addition of Defeat Branch says that adding a defeat branch to any argument degrades its ranking:

If $AF' = AF \cup P_-(A)$ then $A_{AF} >_d A_{AF'}$

This does not hold either. A counterexample is displayed in the bottom-left part of Figure 15. For both AF and AF' we have that $A \in \forall$ for all four classical semantics. If, moreover, all arguments have undefeated defeaters in *UAF*, then A has, for all n, the same n-defeat points in AF and in AF', namely, $\{A\}$ and $\{C\}$. So then $A_{AF} \approx^n_d A_{AF'}$. Note that D is irrelevant to the status of A given that C already defends A against B, so $\{D\}$ is not an n-defeat point of A.

Increase of Defeat Branch says that increasing the length of a defeat branch of an argument improves its ranking:

If $B \in BR^-(A), B \notin BR^+(A)$ and $AF' = AF \cup P_+(B)$ then $A_{AF'} >_d A_{AF}$.

This does not hold. For a counterexample, consider $AF = A \leftarrow B$ and assume that AF can be expanded to AF' by adding $B \leftarrow C \leftarrow D$ after which there are no further expansions. Then in all four classical semantics $A \in \forall^\neg$ for both AF and AF', so $A_{AF} \approx^n_d A_{AF'}$ for all n.

Increase of Defense Branch says that increasing the length of a defense branch of an argument degrades its ranking:

If $B \in BR^+(A), B \notin BR^-(A)$ and $AF' = AF \cup P_+(B)$ then $A_{AF} >_d A_{AF'}$.

This does not hold. A counterexamples is displayed in the bottom-right part of Figure 15. Assume that E has no defeaters in *UAF*. Then no new defeat points for A are created so $A_{AF} \approx^n_d A_{AF'}$ for all n.

Finally, **Total**, which says that all pairs of arguments can be compared, was above shown to hold for the special case where $\geq_c$ is total (Proposition 16). Figure 9 illustrates that $\geq_c$ is not in general total; then Proposition 16 implies that, for all n, $\geq_d^n$ is not in general total.

Why do most principles fail to hold? This is for several reasons. One reason is that several principles have been shown to be incompatible [11]: Cardinality Precedence is incompatible with Quality Precedence, with Defeat vs Full Defense and with Addition of Defense Branch, while Void Precedence is incompatible with Strict Addition of Defense Branch. Apart from this, some principles 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. Some principles (also) 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. Finally, some principles fail since they in one way or another compare numbers or sets of defeaters of arguments, while our definition of dialectical strength instead looks at the points at which an argument can be defeated, regardless of how many defeaters it actually has. As further explained below in Section 8, this is for good reasons. These observations suggest that with the classical contextual ordering there is not much room for adjusting these principles in ways that fit with our notion of dialectical strength.

4.4 Contextual Strength with Burden Semantics

It is interesting to study cases where Definition 20 is combined with some ranking-based semantics proposed in the literature as the basis x for the contextual ordering $\geq_c^x$. This illustrates the generality of our approach but it may also shed light on the extent to which failure of principles is due to the extension-based nature of Dung's classical semantics.

First, the following proposition is implied by Proposition 14 and the definition of the mentioned properties.

Proposition 27 The following principles are, for any contextual semantics x and for all n, satisfied by $\geq_d^{x,n}$ if they are satisfied by $\geq_c^x$: Void Precedence, Cardinality Precedence, Quality Precedence, Defense Precedence, Strict Counter-Transitivity, Distributed Defense Precedence, Defeat vs Full Defense, Strict Addition of Defense Branch, Addition of Defense Branch, Addition of Defeat Branch, Increase of Defeat Branch, Increase of Defense Branch.

PROOF. This holds since all these principles require a $>_c^x$ relation to hold under specified conditions. Then Proposition 14 implies that the corresponding $>_d^{x,n}$ relation holds under the same conditions. QED

Let us by way of example consider the ranking-based semantics that satisfies the highest number of principles discussed in [11], namely, [2]'s *burden semantics*. According to [11], burden semantics satisfies all principles except Self-contradiction, Quality Precedence, Defeat vs Full Defense, Addition of Defense Branch and Strict Addition of Defense Branch. Therefore, counterexamples to these 5 principles for dialectical strength with burden semantics can be constructed by assuming that no expansion is allowed, or that no expansion exists that defeats any argument in a given AF. Furthermore, Proposition 27 implies that dialectical strength with burden semantics satisfies

Cardinality Precedence, Defense Precedence, Strict Counter-Transitivity, Distributed Defense Precedence, Addition of Defeat Branch, Increase of Defeat Branch and Increase of Defense Branch. Moreover, Proposition 16 implies that dialectical strength with burden semantics satisfies Total. Finally, our new postulate Weak Defense Precedence is satisfied since it is implied by Defense Precedence, which is satisfied by burden semantics. It is left to investigate Abstraction, Independence, Counter-Transitivity and Non-defeated Equivalence.

We first repeat the definitions for burden semantics as presented in [11], adapted to the present notational conventions. The idea of burden semantics is that it recursively assigns a burden number to arguments based on the burden numbers of its defeaters. This is formalised as a step-wise process.

Definition 23 Let $AF = (\mathcal{A}, \mathcal{D})$ be an abstract argumentation framework, $A \in \mathcal{A}$ and $i \in \mathbb{N}$.

$$Bur_i(A) = \begin{cases} 1 & \textit{if } i = 0; \\ 1 + \Sigma_{B \in (A^-)} \frac{1}{Bur_{i-1}(B)} & \textit{otherwise} \end{cases}$$

Here $\Sigma_{B \in A^-} \frac{1}{Bur_{i-1}(B)} = 0$ by convention if $A^- = \emptyset$. The *Burden number* of A is denoted $Bur(A) = \langle Bur_0(A), Bur_1(A), \ldots \rangle$.

Two arguments are then lexicographically compared on the basis of their Burden numbers.

Definition 24 For any $AF = (\mathcal{A}, \mathcal{D})$ and all $A, B \in \mathcal{A}$ it holds that $A \geq_c^b B$ iff $Bur(B) \geq_{Lex} Bur(A)$.

In step 1 the definition counts the number of defeaters of each argument (and increases the number with 1). The higher this number, the worse. Then in step 2 these numbers are lowered to the extent that the argument's defeaters themselves have defeaters, and so on. The following example, taken from [2], illustrates burden semantics.

Example 28 Let $AF = A \leftarrow B \leftarrow C$. Then the burden numbers are as follows:

	A	B	C
step 1	2	2	1
step 2	1.5	2	1
.	.	.	.

The lexicographic ordering on these burden numbers yield that $A <_c^b B$, unlike in the classic contextual ordering, where A and C are of equal rank, whatever semantics is adopted.

The next example (also taken from [2]), shows that there are no clear relations between burden semantics and the classic contextual ordering and that they in fact behave very differently.

Example 29 In the AF in Figure 16 we have the following burden numbers:

	A	B	C	D	E	F	G
step 1	3	2	2	1	1	1	2
step 2	2	2	2	1	1	1	2
.	.	.	.	.	.	.	.

Burden semantics gives $A <_c^b G$ (since A has more defeaters than G) while the classic contextual ordering instead yields $G <_c A$ whatever semantics is adopted, since A is successfully defended against all its defeaters while G is not defended against F.

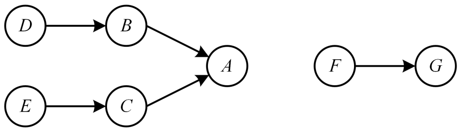

Figure 16: Burden semantics v. classical contextual ordering

For **Abstraction** it is easy to see that the counterexample given in Section 4.3 for classic contextual strength also holds for burden semantics as the contextual ordering, since Burden semantics satisfies Abstraction.

Recall that **Independence** says that the ranking between two arguments in the same AF should be independent of arguments that are connected to neither. This does not hold. The counterexample for all four classic semantics in Figure 13 also holds here. Recall that $cc(AF) = \{AF', AF''\}$ where $AF' = A \leftrightarrow B$ while $AF'' = C$. Then we have $A_{AF'} \approx^b_c B_{AF'}$ and $A_{AF} \approx^b_c B_{AF}$ since they defeat each other and have no other defeaters in AF' or AF. So then we have to look at their defeat points just as in Figure 13, so we again have $A_{AF'} \approx^n_d B_{AF'}$ but $B_{AF} >^n_d A_{AF}$ for all n.

Recall that **Counter-Transitivity** says that if the direct defeaters of B are at least as numerous and strong as those of A in the same AF, then A is at least as strong as B. This does not hold. A counterexample is displayed in Figure 17, where we assume that A has one attack target while B has two, and that both of them have undefeated defeaters in *UAF* on all their attack targets. It is easy to see that A and B have the same burden number so $A \approx^b_c B$. However, A has, for all n, one defeat point while B has two, so $A >^n_d B$.

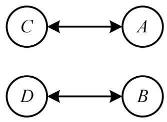

Figure 17: Counterexample to Counter-Transitivity with Burden semantics

Finally, recall that **Non-Defeated Equivalence** says that all non-defeated arguments are of equal strength. It is easy to see that the above counterexample to Abstraction is still a counterexample to Non-defeated equivalence.

This analysis is summarised in the following proposition.

Proposition 30 When combined with burden semantics as contextual strength $\geq^b_c$, Definition 20 of dialectical argument strength satisfies Void Precedence, Cardinality Precedence, Defense Precedence, Strict Counter-Transitivity, Distributed Defense Precedence, Total, Addition of Defeat Branch, Increase of of Defeat Branch and Increase of Defense Branch while it violates Abstraction, Independence, Counter-Transitivity, Self-contradiction, Quality Precedence, Defeat vs Full Defense, Strict Addition of Defense Branch, Addition of Defense Branch and Non-Defeated Equivalence.

Table 1 summarises the results of Sections 4.2, 4.3 and 4.4. Since the results, respectively, counterexamples hold for all n, there is no need to refer to n in the table. While

Principle	Ranking with classical	Ranking with burden
Abstraction	×	×
Independence	×	×
Void Precedence	×	✓
Self-Contradiction	×	×
Cardinality Precedence	×	✓
Quality Precedence	s	×
Counter-Transitivity	×	×
Strict Counter-Transitivity	×	✓
Defense Precedence	×	✓
Weak Defense Precedence	✓	✓
Distributed-Defense Precedence	s	✓
Strict Addition of Defense Branch	×	×
Addition of Defense Branch	×	×
Increase of Defeat Branch	×	✓
Addition of Defeat Branch	×	✓
Increase of Defense Branch	×	✓
Total	s	✓
Non-Defeated Equivalence	×	×
Defeat vs Full Defense	✓	×

Table 1: Overview of results for classical and burden semantics. A ✓indicates that the principle is satisfied in general, an 's' indicates that it does not hold in general but that we have identified a special case in which it holds, and a $\times$ indicates that it does not hold in general and that we have not identified a special case in which it holds, although such special cases may exist.

these are more positive results than with classic contextual strength, for all the principles that dialectical strength with burden semantics satisfies except Total, the reason why it satisfies the principle is that strict contextual preference implies strict dialectical preference (Proposition 14). Does this mean that burden semantics is a better choice for our purpose than classical contextual strength? We believe not, since as shown by Example 29, burden semantics does not properly model the effect of defending arguments against defeaters, which we believe is an essential feature of argumentation, so beautifully captured in Dung's semantics of abstract argumentation frameworks. Furthermore, we have seen that with burden semantics as contextual strength, $\geq_d^{b,n}$ violates, for all n, four principles in addition to those violated by burden semantics as $\geq_c^b$, namely, Abstraction, Independence, Counter-Transitivity and Non-defeated Equivalence. In conclusion, we can say that our negative results can only to a limited extent be attributed to the choice of classic instead of burden semantics as contextual strength.

5 Instantiating the Abstract Notions of Expansions with *ASPIC*$^+$

We next study how the abstract framework of the previous sections can be instantiated with *ASPIC*$^+$. This first requires an instantiation of the abstract definitions of (allowed) expansions of Section 3.2 (Definitions 14 and 15). This in turn requires a specification of how the *UAF* can be generated by a universal *structured* argumentation framework *USAF* to which it corresponds. Since an *SAF* is in *ASPIC*$^+$ determined by an argumentation theory, we must also specify the notion of a universal argumentation theory *UAT*. Figure 18 visualises how we have refined the abstract theory of expansions in Definition 15 (on the left, repeated bottom right) and how we now will instantiate this definition with structured expansions (on the right). On the right, the only things that are nontrivial to define are what is a *UAT*, how to go from a *UAT* to an AT and how to expand an AT into an AT'. The rest follows from the way *ASPIC*$^+$ is defined (the black arrows follow directly, the blue ones by straightforward counterparts of existing definitions).

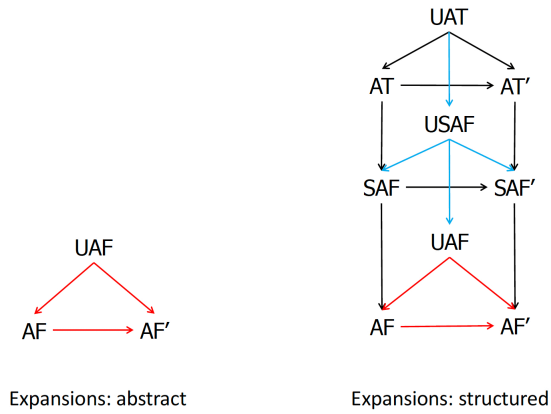

Figure 18: Expansions: abstract versus structured

5.1 Universal Structured Argumentation Frameworks

A *UAF* is now defined as corresponding to a universal structured argumentation framework, which is in turn defined by a universal argumentation theory. Together, they define the space of possible knowledge bases, possible sets of inference rules and possible argument orderings and thus define the space of possible argumentation frameworks.

Definition 25 **[Universal Argumentation theories and universal structured *AFs*]** A *universal argumentation theory* is a tuple $UAT = ((\mathcal{L}^u, \mathcal{R}^u_s \cup \mathcal{R}^u_d, n^u), \mathcal{K}^u_n \cup \mathcal{K}^u_p)$ where all elements are defined as for *ASPIC*$^+$ argumentation theories except that $\mathcal{K}^u_n$ and $\mathcal{K}^u_p$ do not have to be disjoint. Then a *universal structured argumentation framework* defined by UAT is a tuple $USAF = (\mathcal{A}^u, \mathcal{C}^u, \preceq^u)$ defined according to Definition 10, where $\preceq^u$ is some preference ordering on $\mathcal{A}^u$. A $UAF = (\mathcal{A}^u, \mathcal{D}^u)$ that is the abstract argumentation framework corresponding to some given *USAF* is denoted by *sUAF*.

Example 31 Consider a *UAT* with

$\mathcal{L}^u = \{p, \neg p, q, \neg q, r, \neg r, d, \neg d, d', \neg d'\}$,
$\mathcal{R}^u_s = \emptyset$,
$\mathcal{R}^u_d = \{p \Rightarrow q; \neg r \Rightarrow \neg q\}$,
$n^u = \{(p \Rightarrow q, d), (\neg r \Rightarrow \neg q, d')\}$,
$\mathcal{K}^u_n = \{p\}$,
$\mathcal{K}^u_p = \mathcal{L}^u$

The arguments and direct attack relations of the *USAF* defined by this *UAT* are visualised in the top part of Figure 19, in which the argument names are placed next to the argument's conclusion. Assuming the empty argument ordering, all attacks succeed as defeats. Then the corresponding *UAF* is visualised in the bottom part of the figure, which also displays the indirect defeat relations.

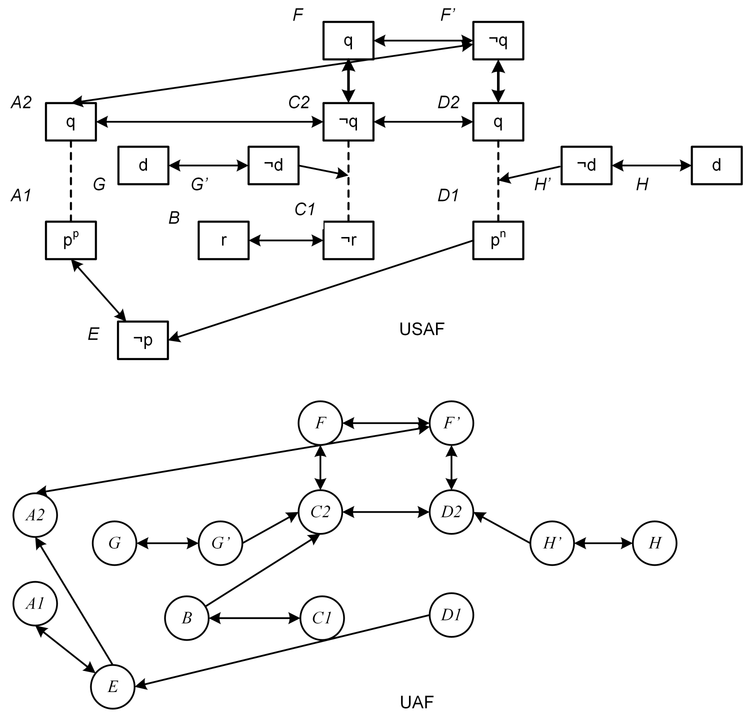

Figure 19: *USAF* defined by *UAT* and corresponding *UAF*

Note that the sets $\mathcal{R}_s^u$ and $\mathcal{R}_d^u$ of a UAT are not required to contain all well-formed strict, respectively, defeasible rules over $\mathcal{L}^u$. This is to allow for instantiations where the strict rules are defined by a logical interpretation of $\mathcal{L}^u$ and/or the defeasible rules correspond to some set of argument schemes. The limiting case where $\mathcal{R}_s^u$ and $\mathcal{R}_d^u$ do contain all well-formed rules over $\mathcal{L}$ is suitable for applications where the choice of strict and/or defeasible rules is fully free, as in online debate settings. For similar reasons $\mathcal{K}_p^u$ and $\mathcal{K}_p^u$ are not required but are allowed to equal $\mathcal{L}^u$. The reason why $\mathcal{K}_n^u$ and $\mathcal{K}_p^u$ can overlap is to allow that the type of a premise is unspecified until determined when constructing an AT in UAT. Accordingly, to keep the notion of an argument on the basis of a UAT well-defined, we now assume that in Definition 8(1) it is explicitly indicated whether a premise is taken from $\mathcal{K}_n^u$ or from $\mathcal{K}_p^u$.

The structural counterpart of Definition 14 is as follows.

Definition 26 [**Argumentation theories and structured *AFs* in a universal *AT***] An *argumentation theory in* a given UAT is an *ASPIC*$^+$ argumentation theory $AT = ((\mathcal{L}, \mathcal{R}_s \cup \mathcal{R}_d, n), \mathcal{K}_n \cup \mathcal{K}_p)$ where

- $\mathcal{L} \subseteq \mathcal{L}^u$;

- $\mathcal{R} \subseteq \{S \rightsquigarrow \varphi \in \mathcal{R}^u \mid S \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}\}$;
- $n = n^u \cap \{(r, \varphi) \mid r \in \mathcal{R}_d\}$;
- $\mathcal{K}_n \subseteq \mathcal{K}_n^u$;
- $\mathcal{K}_p \subseteq \mathcal{K}_p^u$.

A *structured argumentation framework in* UAT is a structured argumentation framework $SAF = (\mathcal{A}, \mathcal{C}, \preceq)$ defined by an AT in UAT for some ordering $\preceq$ on $\mathcal{A}$ such that for the $AF = (\mathcal{A}, \mathcal{D})$ corresponding to SAF it holds that $\mathcal{D} \subseteq \mathcal{D}^u$.

Example 31 (Cont.) Consider the following AT in the above UAT:

$\mathcal{L} = \{p, \neg p, q, \neg q, r, \neg r, d\}$,
$\mathcal{R}_s = \emptyset$,
$\mathcal{R}_d = \mathcal{R}_d^u$,
$n = n^u$,
$\mathcal{K}_n = \emptyset$,
$\mathcal{K}_p = \{p, r\}$

Combined with any argument ordering, the *SAF* defined by this *AT* contains three arguments:

A_1: p $\quad$ A_2: $A_1 \Rightarrow q$ $\quad$ B: r

and no attack relations; see also Figure 20 below on the left. The corresponding *AF* equals $(\{A_1, A_2, B\}, \emptyset)$.

In our approach, the constraints of Definition 26 are the minimal constraints on formulating ATs in a UAT. If desired, further constraints can be imposed in the definition of allowed expansions in terms of preserving a given type of AT (see Definition 27 below). Some possible *types of* AT in a UAT can be defined as follows.

An argumentation theory in UAT is

- *objective* iff $\mathcal{K}_n = \mathcal{K}_n^u \cap \mathcal{L}$;
- *logic-based* iff $\mathcal{R}_s = \{S \to \varphi \in \mathcal{R}_s^u \mid S \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}\}$;
- *strongly logic-based* iff $\mathcal{R} = \{S \rightsquigarrow \varphi \in \mathcal{R}^u \mid S \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}\}$.

Clearly, every strongly logic-based AT is logic-based.

Example 31 (Cont.) AT is trivially strongly logic-based since it has no strict rules, but it is not objective.

Objective ATs are called objective since they accept all necessary premisses from UAT that can be expressed in their language. Objective ATs may be suitable for knowledge-based systems (such as for medical diagnosis or crime investigation), in which the general knowledge is fixed but investigations must be done to gather specific observations (such as medical tests on a person who is ill, or searching for evidence predicted by a crime scenario). (Strongly) logical ATs are called thus since they accept all (defeasible and) strict inference rules from UAT that can be expressed over their language. This feature can be used to encode a logic (in $\mathcal{R}_s^u$) or a recognised set of argument schemes (in $\mathcal{R}_d^u$). Consider, for example, a UAT with $\mathcal{L}^u$ a propositional language and $\mathcal{R}_s^u = \{S \to \varphi \mid S \subseteq \mathcal{L}^u$ and S is finite and $\varphi \in \mathcal{L}^u$ and $S \vdash \varphi\}$ where $\vdash$ denotes propositional-logical consequence. Then all logic-based ATs in UAT allow for deductive reasoning with the full power of propositional logic over their language. Further constraints on ATs can be formulated. For instance, they could be required to be well-formed in the sense of [49], who impose restrictions on the type of argument ordering, the consistency of $\mathcal{K}_n$ and the syntax of $\mathcal{R}_s$ in order to make the system satisfy consistency and closure postulates. Or if the UAT has the language restriction of assumption-based argumentation, then ATs can, if desired, be required to be 'flat' in that an assumption cannot appear in the consequent of a strict rule [69].

The following proposition captures that Definitions 25 and 26 indeed instantiate Definitions 14 and 15 since it implies that every AF that can be generated from a universal argumentation theory is an AF in the same universal AF as required by Definition 14 (which is used in Definition 15).

Proposition 32 Given a $USAF = (\mathcal{A}^u, \mathcal{C}^u, \preceq^u)$ in a $UAT = ((\mathcal{L}^u, \mathcal{R}^u_s \cup \mathcal{R}^u_d, n^u), \mathcal{K}^u_n \cup \mathcal{K}^u_p)$, an $AF = (\mathcal{A}, \mathcal{D})$ corresponding to an $SAF = (\mathcal{A}, \mathcal{C}, \preceq)$ in UAT is an AF in $sUAF = (\mathcal{A}^u, \mathcal{D}^u)$, where $sUAF$ corresponds to $USAF$.

PROOF. It holds that $\mathcal{A} \subseteq \mathcal{A}^u$ by definition of an SAF in UAT. Furthermore, by the same definition it holds that $\mathcal{D} \subseteq \mathcal{D}^u$, so no additional defeat relations are possible. QED

5.2 Allowed Expansions

So far all we have done is instantiating the notion of an AF in a *UAF* for *ASPIC*$^+$(as captured by Proposition 32). The next step is to define the *allowed expansions* of an AF that corresponds to an SAF in a universal argumentation theory. The main task is to ensure that the result of such an expansion still corresponds to a structured AF in the universal argumentation theory, in order to respect the structural constraints imposed by *ASPIC*$^+$. A natural way to achieve this is to require that an expansion of AF is the result of additions to the AT that defines the SAF to which the AF corresponds: the extended AT will define a new SAF that will in terun induce a corresponding expanded AF, as illustrated in Figure 18. This is directly stated by the following definition. It assumes that the argument ordering $\preceq$ of an SAF comes with a definition of its type, as, for example, the definitions of a basic, weakest- or last link ordering [51]. Likewise, it assumes a definition of an allowed type of AT (see the above discussion of types of AT in Section 5.1). Note that this definition only captures the structural constraints following from the *ASPIC*$^+$ definitions; if required by the nature of the application, further constraints on allowed expansions can be added.

Definition 27 [**Allowed expansions**] Consider any AF in a given *sUAF* that corresponds to an $SAF = (\mathcal{A}, \mathcal{C}, \preceq)$ in UAT defined by $AT = ((\mathcal{L}, \mathcal{R}, n), \mathcal{K})$, and consider any AF' in *sUAF* that expands AF. Then AF' is an *allowed expansion* of AF given *UAF* iff AF' corresponds to an $SAF' = (\mathcal{A}', \mathcal{C}', \preceq')$ in UAT defined by $AT' = ((\mathcal{L}', \mathcal{R}', n'), \mathcal{K}')$ such that:

1. $\mathcal{L} \subseteq \mathcal{L}'$;
2. $\mathcal{R} \subseteq \mathcal{R}'$;
3. $\mathcal{K}_n \subseteq \mathcal{K}'_n$ and $\mathcal{K}_p \subseteq \mathcal{K}'_p$;
4. for $\preceq'$ it holds that
   (a) $\preceq \subseteq \preceq'$;
   (b) $A \prec' B$ if $A \prec B$;
5. AT' is of the same type as AT.

Note that the constraints on $\preceq'$ together make that SAF' *extends* SAF in the sense of Modgil and Prakken [48]. In this case we also say that $\preceq'$ extends $\preceq$. The constraints ensure that symmetric defeats can be 'resolved' into asymmetric defeats but that asymmetric defeats remain untouched.

Example 31 (Cont.) Suppose that someone wants to extend AT in a way that makes B overruled. Then given *UAT* this can only be done by adding r to $\mathcal{K}_p$ and adding $(r, \neg r)$ to $\preceq$. This results in

$$\mathcal{L}' = \mathcal{L}, \mathcal{R}'_s = \mathcal{R}_s, \mathcal{R}'_d = \mathcal{R}_d, n' = n, \mathcal{K}'_n = \mathcal{K}_n,$$
$$\mathcal{K}'_p = \{p, q, \neg r\}$$
$$\preceq' = \preceq \cup \{(r, \neg r)\}$$

The arguments and direct attacks in the *SAF'* defined by AT' are visualised in Figure 20 on the right. Combined with $\preceq'$ the corresponding AF' is as visualised on the right of

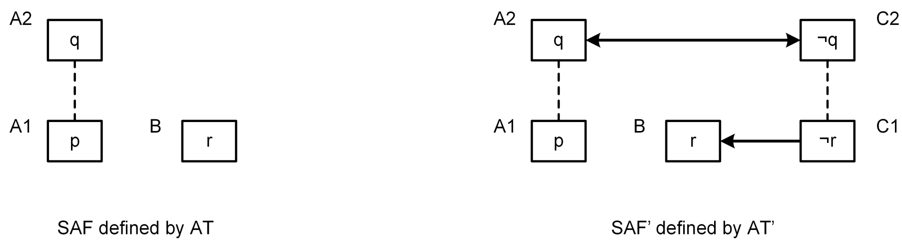

Figure 20: *SAF* defined by *AT* and *SAF*' defined by *AT*'

the figure. As regards the corresponding A's, the AF corresponding to AT is visualised on the left of the Figure 21. At first sight, it would seem that the addition of r to AT

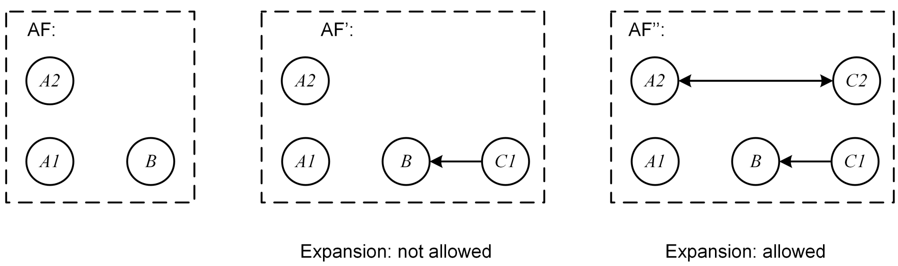

Figure 21: Allowed and unallowed abstract expansions

at the abstract level results in expanding with only argument $C1$ from Figure 19 and a single defeat relation from C_1 to B. This would yield AF' in Figure 21. However, this expansion is unallowed: since the rule $\neg r \Rightarrow \neg q$ is in AT', argument C_2 must also be added. Moreover, on the basis of $\preceq'$, which contains no preference between C_2 and A_2, a mutual defeat relation between A_2 and C_2 has to be added. The result is AF'' in Figure 21. Assuming $\preceq$ and $\preceq'$ are of the same type, it is easy to see that AF'' is an allowed expansion of AF. Note that if no preference was added, then the defeat relation between B and C_1 would also be mutual while, moreover, a defeat relation from B to C_2 would have to be added. Example 31 further illustrates that purely abstract accounts of expansions like [8] implicitly make assumptions that are not in general satisfied by instantiations.

5.3 Defining Attack Targets for *ASPIC*$^+$ Arguments

Finally, an argument's **attack targets** (cf. Definition 17) must be defined. This paper adapts a definition of [38] (who did not embed the definition in a setting with expansions as in this paper). All subarguments on which an argument can be attacked are attack targets but an issue is whether a distinction should be made between targets of undercutting and rebutting attack. Such a distinction seems reasonable since arguing that a rule does not apply to the case at hand is a fundamentally different kind of attack than arguing that there are reasons against its consequent [38]. Hence:

Definition 28 Given a $sUAF = (\mathcal{A}^u, \mathcal{D}^u)$, for any $A \in \mathcal{A}^u$ we define $t(A)$ as $\texttt{Prem}_p(A) \cup \{(B, \varphi) \mid B \in \texttt{Sub}(A) \text{ and } \varphi = \texttt{Conc}(B) \text{ and } \texttt{TopRule}(B) \in \mathcal{R}_d\} \cup \{(B, r) \mid B \in \texttt{Sub}(A) \text{ and } r = \texttt{TopRule}(B) \text{ and } r \in \mathcal{R}_d\}$.

Whenever B defeats A, we say that B defeats A *on* $t \in t(A)$ iff B successfully undermines or rebuts A on t or B undercuts A on t.

It holds that the number of attack targets of an argument equates the number of its ordinary premises plus twice its number of subarguments with a defeasible top rule.

Property: $\mid t(A) \mid \ = \ \mid \texttt{Prem}_p(A) \mid + 2 \times \mid \{A' \in \texttt{Sub}(A) \mid \texttt{TopRule}(A') \in \mathcal{R}_d\} \mid$.

We can now state that Definition 18 is well-defined for the *ASPIC*$^+$ setting in that a set S' as required in that definition always exists (recall that such existence is not relative to a *UAF*). A set S' can always be constructed since for each attack target t of an *ASPIC*$^+$ argument an argument $-t$ can be constructed where $-t$ is assumed to be in $\mathcal{K}_n$.

5.4 Examples: Dialectical versus Rhetorical Strength

We next discuss two realistic examples that also illustrate that it is good to have separate accounts of dialectical and rhetorical strength. First, consider the following two natural-language examples.

A: You are not allowed to do a retake since you missed too many lectures.
B: You are not allowed to do a retake since you missed too many lectures, so you made insufficient effort.

We formalise this in an AT with $R_s = \emptyset$, $\mathcal{R}_d = \{r_1, r_2, r_3\}$ where

r_1: *missed* $\Rightarrow \neg$ *retake*
r_2: *missed* $\Rightarrow \neg$ *effort*
r_3: $\neg$ *effort* $\Rightarrow \neg$ *retake*

and with $\mathcal{K}_n = \{\textit{missed}\}$ and $\mathcal{K}_p = \emptyset$. The *SAF* defined by AT consists of the arguments A_1, A_2, B_2, and B_3 displayed in the top right part of Figure 22, which have no attack relations. Then with any $\preceq$ the corresponding $AF = (\{A_1, A_2, B_2, B_3\}, \emptyset)$, shown in the bottom-left part of Figure 22. Focusing on A_2 and B_3, both are sceptically acceptable in any classical semantics, so $A2_{AF} \approx_c B3_{AF}$, while they have the following attack targets:

$t(A_2) = \{(A_2, \neg \textit{retake}), (A_2, r_1)\}$
$t(B_3) = \{(B_2, \neg \textit{effort}), (B_2, r_2), (B_3, \neg \textit{retake}), (B_3, r_3)\}$.

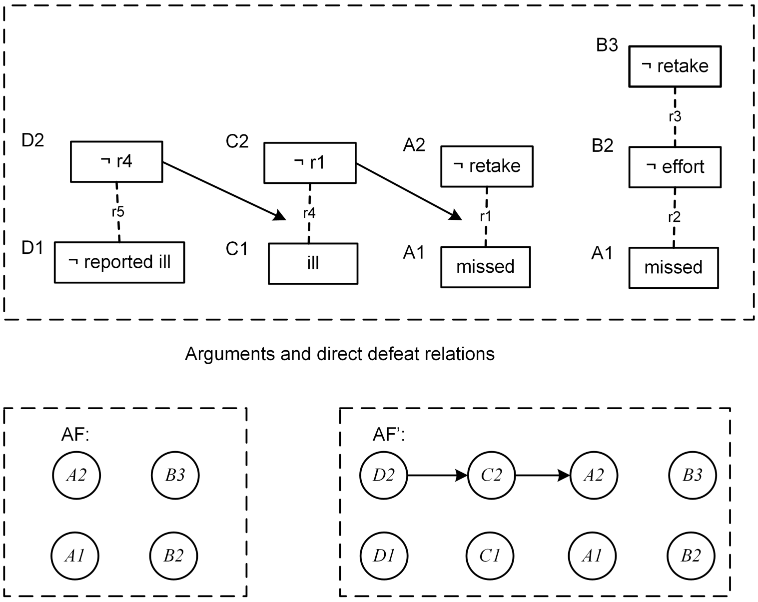

Figure 22: Dialectical versus rhetorical strength

So $|t(A_2)| < t|(B_3)|$. Assume furthermore that *UAF* (not shown in Figure 22) contains undefeated defeaters on all these attack targets which, moreover, have no defeat relations with any other argument. Then for all n the sets of n-defeat points of A, respectively, B, consist of, respectively, all their attack targets as singleton sets. So, applying Definition 20, it holds that $A_2 >_d^{x,n} B_3$ for all classic x and all n.

Consider next a definition of rhetorical strength of arguments for an audience in the spirit of [39], where the audience is conceived as a set S of rules and premises, where X is at least as persuasive as Y if $(Prem(Y) \cup Rules(Y)) \cap S \subseteq (Prem(X) \cup Rules(X)) \cap S$. That is, persuasiveness is measured in terms of the overlap of the elements of an argument with what the audience believes. Consider then an audience $S = \{missed, r_3\}$. Then the overlap of A_2 with S is $\{missed\}$ while the overlap of 3 with S is $\{missed, r_3\}$, so B_3 is more persuasive than A_2. This illustrates that while sparsely justifying one's claims or decisions may be dialectically good, it may at the same time make an argument less persuasive.

Another illustration of this possibility is the rhetorical principle of *procatalepsis*, which says that speakers can strengthen their argument by dealing with possible counterarguments before the audience can raise them [12]. Consider the following two further arguments.

C: Missing too many lecturers is not a reason for not being allowed to do a retake if one is ill, and I was ill.

D: Although this is true in general, this does not hold if you did not report being ill.

We formalise this by letting AT' add r_4 and r_5 to $\mathcal{R}_d$ where

r_4: *ill* $\Rightarrow \neg r_1$,
r_5: $\neg$ *reported ill* $\Rightarrow \neg r_4$

and letting AT' add *ill* and *reported ill* to $\mathcal{K}_n$. The resulting SAF' is shown in the entire top part of Figure 22, while the corresponding AF' is shown in the bottom-right part of the figure. Assume that all arguments in AF' have undefeated defeaters in *UAF* (not shown). Then A_2 is skeptically acceptable in both AF and AF'. Moreover, A_2 has, for all n, two n-defeat points in AF, namely, $\{(A_2, \neg \textit{retake})\}$ and $\{(A_2, r_1)\}$, while A_2 has two additional n-defeat points in AF', namely, $\{(D_2, \neg r_4)\}$ and $\{(D_2, r_5)\}$. So A_2 is, for all n, dialectically stronger in AF than in AF'. Yet the idea of procatalepsis as formalised in [12] is that A_2 is more persuasive in AF' than in AF, since D_2 in AF' defends A_2 against the objection C_2 that might be raised by the audience.

6 Properties of the Instantiated Model

Since the model of the previous section is an instantiation of the abstract models proposed in this paper, it inherits the properties proven about these abstract models. In this section we investigate which additional properties can be proven because of the specific nature of our instantiation. All results and counterexamples stated in Sections 6.2 and 6.3 hold for all of the four classic contextual semantics.

6.1 General Results on Expansion Games and Dialectical Argument Strength

In Section 3 Propositions 4 and 19 were proven on the condition that all expansions are allowed if and only if they do not omit defeat relations from UAF. Our *ASPIC*$^+$ instantiation does not satisfy this condition for two reasons. First, Definition 27 rules out expansions that do not satisfy the structural constraints of *ASPIC*$^+$, and second, it allows omitting defeats as long as this is done by extending an argument ordering and respecting the structural constraints of *ASPIC*$^+$.

6.2 The Defeat Property

Next, our instantiation of the abstract model with *ASPIC*$^+$ can be shown to satisfy the defeat property in the following sense (see Section 4.1).

Proposition 33 Any AF corresponding to a *SAF* in a *UAT* satisfies the defeat property if attack targets are defined as in Definition 28.

PROOF. Consider an arbitrary AF satisfying the conditions and with arguments $A, B, C \in AF$, and let attack target t be shared by A and B. If $t \in \mathcal{L}$ then t is an ordinary premise of both A and B. But then C defeats A on φ iff C defeats B on φ by definition of defeat by undermining. If t is of the form (A', φ) where $\varphi = \texttt{Conc}(A')$ then C defeats A on A' iff C defeats B on A' by definition of defeat by rebutting. Finally, if t is of the form (A', φ) where $\varphi = \textit{TopRule}(A')$ then C defeats A on A' iff C defeats B on A' by definition of defeat by undercutting. QED

6.3 Principles for Ranking-based Semantics with *ASPIC*$^+$

We next investigate which of the principles for ranking-based semantics discussed in Section 4 that fail in general do hold for our *ASPIC*$^+$ instantiation. Failure in general does not imply failure for our *ASPIC*$^+$ instantiation, since it may be that particular abstract counterexamples cannot be instantiated with *ASPIC*$^+$. Below, when we specify an AF by a set S of arguments only, we leave implicit that AF corresponds to an SAF in an AT that is extended with all elements of the arguments. When a defeat relation requires a preference relation between arguments, we leave that relation implicit. Finally, when relevant we will in the figures indicate the type of premise in a given AF with superscript n or p while we indicate the name of a defeasible rule as a subscript of $\Rightarrow$.

Abstraction still does not hold. The abstract counterexample given in Section 4.3 can be instantiated with $A = p$ and $B = q$ and $C = \neg q$ where $p, \neg q \in \mathcal{K}_n$ while $q \in \mathcal{K}_p$.

Void precedence still does not hold. We instantiate the abstract counterexample of Figure 13 as in the top left of Figure 23.

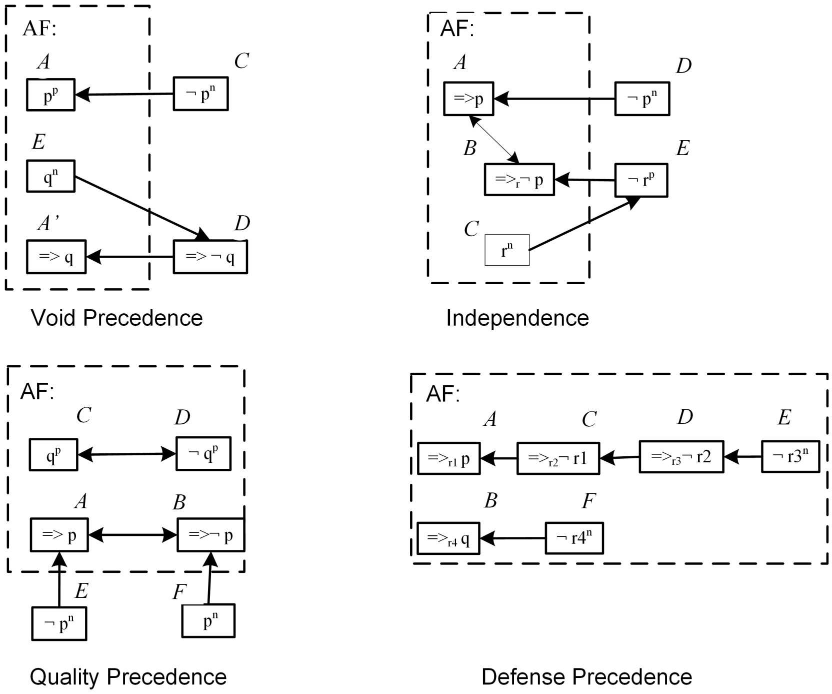

Figure 23: Counterexamples for *ASPIC*$^+$ (1)

Independence still does not hold. We instantiate the abstract counterexample of Section 4.3 as in the top right of Figure 23.

Below we will leave implicit that all given counterexamples instantiate the abstract counterexamples of Section 4.3 except when stated otherwise.

Self-Contradiction still does not hold. Let $A = \ \Rightarrow \neg r$ where $r = n(\Rightarrow_r \neg r)$ for any B and C.

Cardinality Precedence still does not hold. The instantiated counterexample to VP also holds for CP.

Quality Precedence still does not hold. See the bottom left of Figure 23 for an instantiated counterexample.

Defense Precedence still does not hold. See the bottom right of Figure 23 for an instantiated counterexample.

Counter-Transitivity does not hold. A simple instantiated counterexample is an

AF with just A and B where $A = p$ and $B = q$ where *UAF* further only consists of $C = \neg p$ (here $\neg p \in \mathcal{K}_n^u$ while $p, q \in \mathcal{K}_p$).

Strict Counter-Transitivity does not hold since it implies Defense precedence, which does not hold.

We extend and instantiate the abstract counterexample to **Distributed Defense Precedence** as shown in Figure 24. Here we leave the six additional arguments $b, c, d, q, s, t \in$

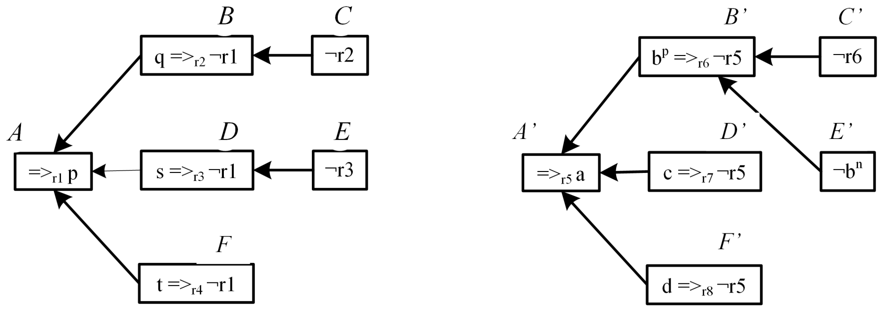

Figure 24: Counterexample to Distributed Defense Precedence

$\mathcal{K}_p$ implicit. Moreover, $\neg b \in \mathcal{K}_n$.

The instantiated counterexamples to Abstraction are also instantiated counterexamples to **Non-Defeated Equivalence**.

Strict Addition of Defense Branch still does not hold. See the top left of Figure 25 for an instantiated counterexample. The only modification of the abstract counterexample is that all arguments in AF do not have one but two attack targets.

The abstract counterexample to **Addition of Defense Branch** can be instantiated as in the top right of Figure 25.

The abstract counterexample to **Addition of Defeat Branch** can be instantiated as in the middle left of Figure 25, where we assume that $B \prec D$.

For a counterexample to **Increase of Defeat Branch**, change E in the counterexamples to Addition of Defeat Branch as in the bottom right of Figure 25 and add F and G as in the figure.

Finally, for a counterexamples to **Increase of Defense Branch**, add D and E to the counterexample to Strict Addition of Defense Branch as in the bottom left of Figure 25.

In conclusion, apart from the fact that the condition of Proposition 21 is satisfied, no new positive results are obtained for the instantiation of our ranking-based semantics with *ASPIC*$^+$.

7 Related Research

In this section we discuss several strands of related research.

7.1 Related Research on Aspects of Argument Strength

Throughout this paper we have made detailed comparisons with existing work on principles for models of argument strength. As noted above, most of this work does not indicate which aspect(s) of argument strength is or are modelled. A recent exception is [12], who model two aspects of "persuasiveness", i.e., of rhetorical strength. The

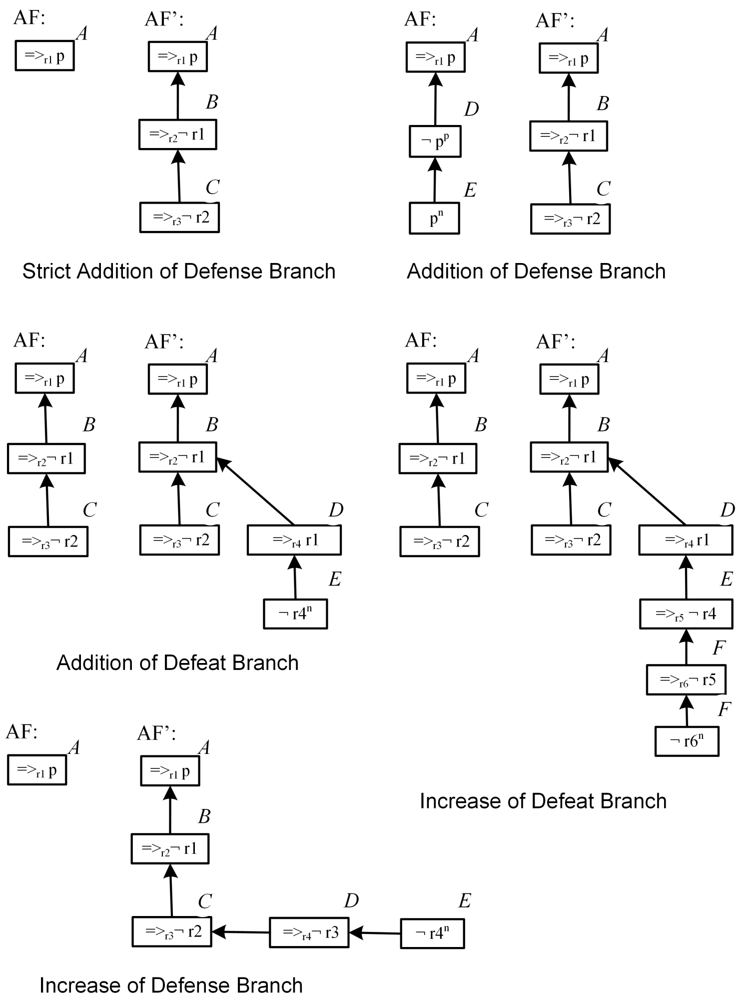

Figure 25: Counterexamples for *ASPIC*$^+$ (2)

first is *procatalepsis*, the attempt of a speaker to strengthen their argument by dealing with possible counter-arguments before the audience can raise them (discussed above in Section 5.4). 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. For example, procatalepsis may conflict with the dialectical principle to motivate a decision as sparsely as possible, as illustrated above in Section 5.4.

In the abstract account of [62] not just a ranking-based but also a weighted semantics for dialectical argument strength was developed. In a weighted AF, arguments have an initial numerical weight between 0 and 1 and argument strength is also defined as a number between 0 and 1. In [62] two semantics were shown to be equivalent and the weighted semantics was also instantiated with *ASPIC*$^+$. Then properties proposed in the literature on weighted semantics (in particular those from [5]), were investigated, which similar largely negative results as for ranking-based semantics. In the present paper we have not included a weighted account in order to make the paper not too long and complex. Instead, we leave this for future research.

Recently, Heyninck et al. [35] have studied ranking-based semantics applied to variants of logic-based argumentation without preferences in which arguments can have an inconsistent set of premises. In terms of *ASPIC*$^+$, argumentation is logic-based in their sense if there are no defeasible rules and if R_s corresponds to all valid inferences from finite set according to a Tarskian consequence notion. For this setting Heyninck et al. propose several new postulates. Among other things, they define a property of *monotony* for ranking-based semantics, which in our notation says that if $A, B \in AF$ and $A^- \subseteq B^-$ then $A \geq_d B$. Heyninck et al. remark that monotony is implied by counter-transitivity and by cardinality precedence. Monotony for ranking-based semantics does not hold for our approach, neither for the abstract case nor for the instantiation with *ASPIC*$^+$. The counterexamples to Void Precedence in Sections 3.5 and 6.3 are also counterexamples to the ranking-based versions of Monotony. Heyninck et al. also establish relations between particular classes of ranking semantics for ABFs and so-called *culpability measures*, which quantify the degree of responsibility of a formula in making a set of formulas inconsistent. They observe that this result is only meaningful in settings where arguments can have inconsistent premises. In conclusion, the relevance of Heyninck et al.'s work for the present paper is limited, since it studies a very limited special case of structured argumentation and in particular since it does not distinguish between kinds of argument strength; because of its static setting their approach does not seem meant for dialectical argument strength.

7.2 Related Research on Explanations and Relevance

There has been some work on notions of relevance of arguments for the status of another argument in order to give explanations for that status. Our notion of a relevant set (Definition 18) is conceptually similar but formally unrelated to [29]'s notion of an explanation. This notion is defined in terms of related admissible sets. Given an $AF = (\mathcal{A}, \mathcal{D})$, a set $S \subseteq \mathcal{A}$ is *related admissible* if there exists an $A \in S$ such that S defends A and S is admissible. Then any related admissible set containing A is an *explanation* of A. That the two notions are formally unrelated follows from the following example AF with $A \leftarrow B \leftarrow C$, where we assume that all arguments have undefeated defeaters in the *UAF*. Then $\{A\}$ is relevant to A but is not related admissible since it is not admissible. Moreover, $\{A, C\}$ is related admissible (and a subset-minimal explanation of A) but this set is not relevant to A. Instead the sets relevant to A are $\{A\}$ and $\{C\}$.

Other work in this vein is [14]. One of their notions of relevance that is conceptually close to our notion of a relevant set is that of (sets of) arguments necessary for the acceptance of an argument. This notion is defined relative to a given AF. An argument $B \in AF$ is relevant for the acceptance of to an argument A if, firstly, B is a direct or indirect defender of A and, second, all extensions of AF that contain A also contain B. Although conceptually close, this notion is also formally unrelated to our notion of a relevant set. First, in all of our above examples where $\{A\}$ is relevant to A and does not defeat one of its defeaters, $\{A\}$ is not necessary for the acceptance of A, since A does not defend itself. Second, in Example 8 the set $\{D, G\}$ is necessary for the acceptance of A but it is not relevant to A.

7.3 Related Research on Argumentation Dynamics

To the best of our knowledge there is no earlier formal work that explicitly addresses dialectical argument strength. Arguably, work on enforcing, preserving or realising a particular argument status in dynamic contexts [8, 24, 9] implicitly addresses aspects of dialectical argument strength. In particular, the notion of expansions of argumentation frameworks introduced in [8] is part of our formal model of dialectical argument strength. Compared to the work on enforcement, we are interested in how the contextual strength of an argument can decrease, which can be seen as the opposite of enforcement. Furthermore, while most work on argumentation dynamics remains at the abstract level, we have tried to instantiate our abstract model with a structured account of argumentation, namely, *ASPIC*$^+$. A key ingredient here was our notion of a universal argumentation framework, which defines the space of possible expansions of a given argumentation framework. This notion bears some resemblance to the notion of incomplete argumentation frameworks [9], which will be discussed below.

There is some recent work that respects that there can be structural constraints on argumentation dynamics. Wallner [74] proposes the general notion of constraints for dynamic operations on abstract or dialectical [16] argumentation frameworks, and discusses its application to structured accounts of argumentation. His ideas are motivated by similar considerations as ours, namely, that abstract approaches can make implicit assumptions that are not satisfied by all structured instantiations.

A similar motivation underlies Rapberger and Ulbricht's [64] study of enforcement in assumption-based argumentation (ABA). Rapberger and Ulbricht introduce ABA counterparts of the abstract notion of expansions and enforcement and relate them to a generalisation of abstract AFs called *cvAFs*, in which 'instantiated' arguments x are defined as pairs $(cl(x), vul(x))$ where $cl(x)$ is the argument's *conclusion* while $vul(x)$ is its set of *vulnerabilities*. A *cvAF* is *well-formed* iff for every $x, y \in \mathcal{A}$ it holds that x attacks[7] y iff the conclusion of x equals a vulnerability of y. The authors then prove complexity results and necessary-and-sufficient conditions for enforceability of single arguments in well-formed *cvAFs*. Rapberger and Ulbricht then instantiate *cvAFs* with ABA, where an argument's vulnerabilities are the contraries of its assumptions. This by definition results in well-formed *cvAFs*. They observe that, like for AFs, results for *cvAFs* do not automatically apply to ABA and they separately prove complexity results for enforceability of single ABA arguments.

It is interesting to see how *ASPIC*$^+$ could generate *cvAFs*. For conclusions this is obvious, while the vulnerabilities are the contradictories of all ordinary premises plus the contradictions of all conclusions of any subargument with a defeasible top rule plus the contradictories of all names of defeasible rules used in the argument. Defining attack is then straightforward, namely, A attacks B iff $cl(A) = -v$ for some $v \in vul(B)$. Defining defeat is less straightforward, since the proper application of preferences for determining defeat depends on the structure of arguments, which is lost in a *cvAF* encoding (contrary to Definition 12, which puts the original *ASPIC*$^+$ arguments in an AF). The most sensible way is to record which original *ASPIC*$^+$ argument gave rise to

[7]Rapberger and Ulbricht use 'attack' instead of 'defeat'.

the *cvAF* argument and then define defeat as between the original arguments. Note that the thus generated *cvAFs* are guaranteed to be well-formed with respect to the attack relation but for the defeat relation this is only guaranteed if the argument ordering is empty or simple (the simple ordering says that that $A \preceq B$ iff B is strict-and-firm while A is defeasible or plausible). This is one reason why the approach of Rapberger and Ulbricht is not in general applicable in the present setting. Another reason is that they do not explicitly work with notions like universal (structured) argumentation frameworks. Instead they assume that every potential argument can be added to a *cvAF* and they leave possible restrictions on the space of expansions for future research.

There is other recent structured work on enforcement and related problems. Borg and Bex [13] develop a structured account of enforcement in argumentation dynamics in Borg and Strasser's [15] 'general argumentation setting', in which, among other things, a special case of *ASPIC*$^+$ with no ordinary premises and no preferences was translated. This work is also in part motivated by the observation that abstract approaches can make implicit assumptions that are not satisfied by all structured instantiations. Within the setting of [15], Borg and Bex define several notions of expansions and enforcement. Unlike in our case, these notions of expansions are not formally related to abstract accounts of expansions, nor are they defined in the context of a universal background framework. In particular, unlike in the present paper no characterisation is given of the conditions under which an expansion is allowed. The main focus of Borg and Bex is on enforcement results. They prove such results for several classes of argumentation settings. Most of these results assume that the setting is *contrapositable*, which notion is very similar to the notion in *ASPIC*$^+$ of closure of strict consequence under contraposition (capturing what can be derived from sets of formulas with only strict-rule application). However, unlike in *ASPIC*$^+$, in Borg and Bex's work as applied to Borg and Strasser's [15] translation of *ASPIC*$^+$, contraposition is not restricted to the strict part of *ASPIC*$^+$. Thus most of the results of Borg and Bex only apply to special cases of *ASPIC*$^+$ with no defeasible rules (and no preferences and ordinary premises).

Odekerken et al. [54, 53, 56] study in the context of *ASPIC*$^+$ whether argument and conclusion statuses can change under expansions of the knowledge base, to find out whether searching for further information makes sense. This work is especially motivated by crime investigation scenarios. In this work the set of *future argumentation theories* is defined as the set of all argumentation theories that extend the knowledge base of a given argumentation theory $AT = ((\mathcal{L}, \mathcal{R}, n), \mathcal{K})$ with a subset of a set $\mathcal{Q} \subset \mathcal{L}$ of *queryables*. In [54, 53] no argument ordering is assumed so attack equates defeat, while in [56] the setting is extended with rule preferences and a last-link argument odering. This approach can be reconstructed as an instance of our approach in Section 5 by letting *UAT* be $((\mathcal{L}, \mathcal{R}, n), \mathcal{K} \cup \mathcal{Q})$ and by imposing the further constraint on Definition 27 that an expansion can only add elements to $\mathcal{K}$ and can only take these elements from $\mathcal{Q}$. Formally this makes any (future) argumentation theory strongly logic-based but this is only since the rules capture domain-specific knowledge; no logic is encoded in the rules. Since the knowledge base equates $\mathcal{K}_n$ and can grow, the ATs are not objective.

The work of Odekerken et al. was abstracted by Mailly and Rossit [45] and Odekerken et al. [55] in terms of incomplete argumentation frameworks [9]. Such frameworks divide an AF in a certain and an uncertain part. Incomplete AFs can be 'completed' or 'specified' by making uncertain arguments or attacks certain. An important difference with our notion of universal AFs is that the notion of specification of an incomplete AF assumes that all arguments and attacks are independent of each other, so that there are no constraints on specifying an incomplete AF. An important theme of our paper has been that this assumption is in general not warranted.

It would be interesting to investigate how all this work on argument dynamics can be combined with studies of dialectical argument strength.

7.4 Preliminary Observations on Computational Complexity

Since the present paper's main focus has been on formal modelling, we have chosen to leave a study of complexity and algorithms to future research. Nevertheless, some of the work discussed above in this section allows us to make some preliminary observations on the computational complexity of our abstract ranking-based semantics with the classic contextual ordering (Sections 3 and 4.1). We consider the special case that all expansions are allowed. Clearly, since our approach involves the determination of an argument's contextual status as a key element, all known complexity results for this problem are lower bounds for the problem of determining a relation of dialectical strength between two arguments. See, for instance, [73] for a study in the context of labelling semantics. Another key element for our semantics is finding an expansion of AF that lowers the contextual status of an argument $A \in AF$. This problem can be mapped onto the problem of stability in incomplete argumentation frameworks (*IAFs*) [45, 55]. The combination of the certain and uncertain part of an *IAFs* can be seen as a *UAF* and the certain part of an *IAF* is then an *AF* in a *UAF*. Then for arguments that have maximal contextual strength, that is, that are in $\forall$, the problem of finding a normal expansion of AF that lowers the contextual status of A is equivalent to the problem of IN-stability of [55] (in [9] called the problem of necessary skeptical acceptance of A in the *IAF*) in that a normal expansion that lowers the status of A exists if and only if A is not IN-stable. Actually, since we assume that all expansions are normal, this holds on the condition that all specifications of an *IAF* add at least one argument and do not add new certain relations between old certain arguments. It is known that for grounded semantics IN-stability is CoNP-complete even if all attacks in an *IAF* are certain [55], which result therefore also holds for our problem of finding a normal expansion of AF that lowers the contextual status of A. So this provides a lower bound for our problem of determining a relation of dialectical strength between two arguments. Since the problem of determining whether an argument is in the grounded extension is known to be in P [27], this shows that there are cases where computing dialectical strength is substantially harder than computing contextual strength. Moreover, this lower bound is likely to be loose, among other things since it is known that computing the labelling class of an argument (as is necessary for our notion of contextual strength) is often harder than computing extension membership [73], and also since our model contains other non-trivial computational subproblems, such as determining whether a set of arguments is relevant to a given argument.

8 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 models of argument strength proposed in the literature fail to hold for our model. Moreover, we noted that there seems not much room for adjusting these principles in ways that fit with our approach to dialectical strength. This reveals something about the possible rational foundations of these principles and highlights the importance of distinguishing between kinds of argument strength, something that in the formal literature on argument strength is not always done. This is a first lesson to be drawn from our study. An important aspect of our approach was that we did not only propose an abstract model but also instantiated it with an account of the structure of arguments and the nature of their relations. This was to avoid overly limiting assumptions at the abstract level that may not hold for all instantiations. A key concept here was the notion of an allowed expansion of an argumentation framework, which turned out to be useful for capturing constraints induced by the structured account of argumentation. This was in turn made possible by making expansions relative to a given universal argumentation framework. A second lesson to be drawn from our study is therefore that an important way to validate abstract models of argumentation is to instantiate them with more concrete accounts of argumentation. This is also something that in the formal literature on argument strength is not always done.

Are our partly negative results on satisfaction of the principles proposed in the literature 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. This raises the question how our dialectical semantics should be evaluated in these respects. We believe our semantics does well. It is based on the well-known philosophical distinction between logical, dialectical and rhetorical argument strength and more specifically on [77]'s description of dialectical argument strength. Moreover, our semantics arguably satisfies desirable properties. Proposition 21 says that under reasonable assumptions, if the set of attack targets of argument A is a subset of the set of attack targets of argument B, then A cannot be weaker than B. This captures quite directly the intuition that justifying a decision more sparsely is better. It should be noted that similar results do not hold for cardinality relations between sets of attack targets or defeaters. It would be interesting to investigate whether such relations do hold in particular contexts, perhaps of a probabilistic nature and to be verified experimentally.

Another key feature of our approach is that it does not look at the number or sets of potential or actual defeaters of an argument. A main reason for this is to discourage obstructive behaviour in discussions by launching as many counterarguments as possible, even if they are clearly nonsensical or based on fake facts. Clearly, a rational model of argumentation should not encourage such behaviour. This also justifies our negative results for principles that in one way or another give the actual number of defeaters or subset relations between sets of defeaters an influence on an argument's strength, which are Void Precedence, Cardinality Precedence, Counter-Transitivity, Strict Counter-Transitivity, Non-Defeated Equivalence and Addition of Defeat Branch.

Note also that in general such obstructive behaviour cannot be avoided by imposing constraints on the *UAF*. In many applications the idea of an *UAF* will be that it contains all 'logically' possible arguments, that is, all arguments that are constructible in the given language and from the given set of rules from any knowledge base expressible in the language. This can be generalised even further by allowing any set of rules that can be expressed over the language. Moreover, as soon as the language is infinite, every defeatable argument may have an infinite number of potential defeaters in the *UAF*. Finally, in many contexts poor quality of arguments is supposed to be established in the course of a dialogue instead of to be determined beforehand.

All this is, of course, not a definitive proof that our approach is the right one. One benefit of our proposal is that it is now formulated so that it is open for critical examination and possible improvements or alternatives. This brings us to topics for future research.

In Section 7.4 we already mentioned a study of complexity and algorithms as a topic for future research. Another research topic is to vary elements of the present model. For instance, other ranking-based semantics can be used for contextual strength. Also, the definitions of an expansion game, relevant sets and defeat points could treat undefined relations of contextual strength differently, by allowing, for instance, that a given relation is changed from defined to undefined. As mentioned above in Section 7, it may also be interesting to investigate how research on enforcement, preservation and realisability can be incorporated in notions of dialectical strength, and how a structured account of weighted approaches to argument strength can be developed and compared to the present ranking-based approach.

Our model is flexible in various ways. For instance, AFs can omit defeats from a UAF, not all expansions may be allowed and, above all, the definition of dialectical argument strength is parametrised by the maximum length of an expansion dialogue. These are three reasons why dialectical argument strength cannot simply be defined relative to contextual status in UAF. This is desirable when it comes to applications: of all three ways of flexibility we have argued in our paper that they are needed in many realistic applications. For example, which maximum length of expansion dialogues is appropriate will depend on the nature of the application. However, a downside of this flexibility is that not many strong results can be proven about the general model. An important topic for future research therefore is to investigate properties if particular choices on the various choice points in our approach are made. For example, it would be interesting to investigate whether additional formal properties can be established in contexts where a *UAF* is generated from particular bodies of knowledge or where expansions have to respect particular argument orderings. More generally, embeddings of our model in dialogical or investigative contexts can be studied, in particular, how these embeddings give rise to additional constraints on allowed expansions. Finally, our formal model could be used as a basis for experimental research, to investigate whether properties that do not hold formally can still be verified empirically and probabilistically in particular contexts. Ultimately, we hope that our model and its further developments will aid in building and especially evaluating artificial arguing agents, even those that are not based on formal but on natural-language-processing methods, such as Debater, ChatGPT and their successors.

9 Acknowledgements

We thank the reviewers of our paper for their many very helpful suggestions. We thank one reviewer in particular for suggesting the incorporation of strategic elements, which we have tried to meet with our notion of expansion games. We also thank Daphne Odekerken for her very useful comments on earlier versions of our paper and for her help in matters of complexity.

References

[1] L. Amgoud. A replication study of semantics in argumentation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-19)*, pages 6260–6266, 2019.

[2] L. Amgoud and J. Ben-Naim. Ranking-based semantics for argumentation frameworks. In W. Liu, V.S. Subrahmanian, and J. Wijsen, editors, *Scalable Uncertainty Management. SUM 2013*, number 8078 in Springer Lecture Notes in Computer Science, pages 134–147, Berlin, 2013. Springer Verlag.

[3] L. Amgoud and J. Ben-Naim. Axiomatic foundations of acceptability semantics. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference*, pages 2–11. AAAI Press, 2016.

[4] L. Amgoud, J. Ben-Naim, D. Doder, and S. Vesic. Acceptability semantics for weighted argumentation frameworks. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI-17)*, pages 56–62, 2017.

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

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

[7] P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors. *Handbook of Formal Argumentation*, volume 1. College Publications, London, 2018.

[8] R. Baumann and G. Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 75–86. IOS Press, Amsterdam etc, 2010.

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

[10] F.J. Bex, S.J. Modgil, H. Prakken, and C.A. Reed. On logical specifications of the argument interchange format. *Journal of Logic and Computation*, 23:951–989, 2013.

[11] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. A comparative study of ranking-based semantics for abstract argumentation. In *Proceedings of the 30st AAAI Conference on Artificial Intelligence (AAAI 2016)*, pages 914–920, 2016.

[12] 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.

[13] A. Borg and F.J. Bex. Enforcing sets of formulas in structured argumentation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Eighteenth International Conference*, pages 130–140. IJCAI Organization, 2021.

[14] A. Borg and F.J. Bex. Necessary and sufficient explanations for argumentation-based conclusions. In *Proceedings of the 16th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 21)*, number 12897 in Springer Lecture Notes in AI, pages 45–58, Berlin, 2021. Springer Verlag.

[15] A. Borg and C. Strasser. Relevance in structured argumentation. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18)*, pages 1753–1759, 2018.

[16] 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.

[17] M. Caminada. Contamination in formal argumentation systems. In *Proceedings of the Seventeenth Belgian-Dutch Conference on Artificial Intelligence (BNAIC-05)*, Brussels, Belgium, 2005.

[18] M. Caminada. On the issue of reinstatement in argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence. Proceedings of JELIA 2006*, number 4160 in Springer Lecture Notes in AI, pages 111–123, Berlin, 2006. Springer Verlag.

[19] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171:286–310, 2007.

[20] M. Caminada, S. Modgil, and N. Oren. Preferences and unrestricted rebut. In S. Parsons, N. Oren, C. Reed, and F. Cerutti, editors, *Computational Models of Argument. Proceedings of COMMA 2014*, pages 209–220. IOS Press, Amsterdam etc, 2014.

[21] C. Cayrol and M.-C. Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.

[22] 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.

[23] M. D'Agostino and S. Modgil. A fully rational account of structured argumentation under resource bounds. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI-20)*, pages 1841–1847, 2020.

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

[25] 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.

[26] P.M. Dung and P.M. Thang. Closure and consistency in logic-associated argumentation. *Journal of Artificial Intelligence Research*, 49:79–109, 2014.

[27] W. Dvorak and P.E. Dunne. Computational problems in formal argumentation and their complexity. In P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors, *Handbook of Formal Argumentation*, volume 1, pages 631–687. College Publications, London, 2018.

[28] Y. Elrakaiby, A. Ferrari, P. Spoletini, S. Gnesi, and B. Nuseibeh. Using argumentation to explain ambiguity in requirements elicitation interviews. In *2017 IEEE 25th International Requirements Engineering Conference*, pages 51–60, 2017.

[29] X. Fan and F. Toni. On computing explanations in argumentation. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI 2015)*, pages 1496–1502, 2015.

[30] T.F. Gordon. The Pleadings Game: an exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1993.

[31] N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical-logic arguments: postulates and properties. *Artificial Intelligence*, 175:1479–1497, 2011.

[32] D. Grooters and H. Prakken. Two aspects of relevance in structured argumentation: minimality and paraconsistency. *Journal of Artificial Intelligence Research*, 56:197–245, 2016.

[33] D. Grossi and S. Modgil. On the graded acceptability of arguments in abstract and instantiated argumentation. *Artificial Intelligence*, 275:138–173, 2019.

[34] E. Hadoux and A. Hunter. Comfort or safety? Gathering and using the concerns of a participant for better persuasion. *Argument and Computation*, 10:113–147, 2019.

[35] J. Heyninck, B. Raddaoui, and C. Strasser. Ranking-based argumentation semantics applied to logical argumentation. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI-23)*, pages 3268–3276, 2023.

[36] J. Heyninck and C. Strasser. Revisiting unrestricted rebut and preferences in structured argumentation. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI-17)*, pages 1088–1092, 2017.

[37] M. Hinton and J. Wagemans. How persuasive is AI-generated argumentation? An analysis of the quality of an argumentative text produced by the GPT-3 AI text generator. *Argument and Computation*, 14:59–74, 2023.

[38] L. van Houwelingen. Gradual acceptability for structured argumentation in ASPIC+. Master's thesis, Department of Information and Computing Sciences, Utrecht University, Utrecht, 2022.

[39] A. Hunter. Making arguments more believable. In *Proceedings of the 19th National Conference on Artificial Intelligence*, pages 6269–274, 2004.

[40] A. Hunter. Towards a framework for computational persuasion with applications in behaviour change. *Argument and Computation*, 9:15–40, 2018.

[41] A. Hunter and M. Thimm. Probabilistic reasoning with abstract argumentation frameworks. *Journal of Artificial Intelligence Research*, 59:565–611, 2017.

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

[43] J. Leite and Martins. Social abstract argumentation. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 2287–2292, 2011.

[44] R.P. Loui. Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence*, 14:1–38, 1998.

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

[46] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019.

[47] T. Miller, R. Hoffman, O. Amir, and A. Holzinger, editors. *Artificial Intelligence*, volume 307. 2020. Special issue on Explainable Artificial Intelligence (XAI).

[48] S. Modgil and H. Prakken. Resolutions in structured argumentation. In B. Verheij, S. Woltran, and S. Szeider, editors, *Computational Models of Argument. Proceedings of COMMA 2012*, pages 310–321. IOS Press, Amsterdam etc, 2012.

[49] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.

[50] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation*, 5:31–62, 2014.

[51] 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 286–361. College Publications, London, 2018.

[52] J. Müller and A. Hunter. An argumentation-based approach for decision making. In *Proceedings of the 25th International Conference on Tools with Artificial Intelligence (ICTAI 2013)*, pages 564–571, 2012.

[53] D. Odekerken, F.J. Bex, A. M. Borg, and B. Testerink. Approximating stability for applied argument-based inquiry. *Intelligent Systems with Applications*, 16:200110, 2022.

[54] D. Odekerken, A. M. Borg, and F.J. Bex. Estimating stability for efficient argument-based inquiry. In H. Prakken, S. Bistarelli, F. Santini, and C. Taticchi, editors, *Computational Models of Argument. Proceedings of COMMA 2020*, pages 307–318. IOS Press, Amsterdam etc, 2020.

[55] D. Odekerken, A. M. Borg, and F.J. Bex. Justification, stability and relevance in incomplete argumentation frameworks. *Argument and Computation*, Pre-press:1–58, 2023.

[56] D. Odekerken, T. Lehtonen, A.M. Borg, J.P. Wallner, and M. Järvisalo. Argumentative reasoning in *ASPIC*$^{+}$ under incomplete information. In *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning*, pages 531–541. IJCAI Organization, 2023.

[57] J.L. Pollock. Defeasible reasoning with variable degrees of justification. *Artificial Intelligence*, 133:233–282, 2002.

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

[59] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124, 2010.

[60] H. Prakken. Reconstructing Popov v. Hayashi in a framework for argumentation with structured arguments and Dungean semantics. *Artificial Intelligence and Law*, 20:57–82, 2012.

[61] H. Prakken. Philosophical reflections on argument strength and gradual acceptability. In *Proceedings of the 16th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 21)*, number 12897 in Springer Lecture Notes in AI, pages 144–158, Berlin, 2021. Springer Verlag.

[62] H. Prakken. Formalising an aspect of argument strength: degrees of attackability. In F. Toni et al., editor, *Computational Models of Argument. Proceedings of COMMA 2022*, pages 296–307. IOS Press, Amsterdam etc, 2022.

[63] H. Prakken. Relating abstract and structured accounts of argumentation dynamics: the case of expansions. In *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning*, pages 562–571. IJCAI Organization, 2023.

[64] A. Rapberger and M. Ulbricht. On dynamics in structured argumentation formalisms. *Journal of Artificial Intelligence Research*, 77:563–643, 2023.

[65] N. Rescher and R. Manor. On inference from inconsistent premises. *Journal of Theory and Decision*, 1:179–219, 1970.

[66] M. Schraagen, B. Testerink, D. Odekerken, and F.J. Bex. Argumentation-driven information extraction for online crime reports. In *CKIM 2018 International Workshop on Legal Data Analysis and Mining*, volume 2482 of *CEUR Workshop Proceedings*, 2018.

[67] N. Slonim, Y. Bilu, and C. Alzate. An autonomous debating system. *Nature*, 591:379–384, 2021.

[68] N. Tamani, P. Mosse, M. Croitoru, P. Buche, V. Guillard, C. Guillaume, and N. Gontard. An argumentation system for eco-efficient packaging material selection. *Computers and Electronics in Agriculture*, 113:174–192, 2015.

[69] F. Toni. A tutorial on assumption-based argumentation. *Argument and Computation*, 5:89–117, 2014.

[70] A. Toniolo, T.J. Norman, A. Etuk, F. Cerutti, R. Wentao Ouyang, M. Srivastava, N. Oren, T. Dropps, J.A. Allen, and Paul Sullivan. Supporting reasoning with different types of evidence in intelligence analysis. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'15)*, pages 781–789, 2015.

[71] F.H. van Eemeren, B. Garssen, E.C. Krabbe, A.F. Snoeck Henkemans, B. Verheij, and J.H. Wagemans. *Handbook of Argumentation Theory*. Springer, Dordrecht, 2014.

[72] K. Čyras, A. Rago, E. Albini, P. Baroni, and F. Toni. Argumentative XAI: A survey. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI-21)*, pages 4392–4399, 2021.

[73] Dvorak W. On the complexity of computing the justification status of an argument. In S. Modgil, N. Oren, and F. Toni, editors, *Theory and Applications of Formal Argumentation. First International Workshop, TAFA 2011. Barcelona, Spain, July 16-17, 2011, Revised Selected Papers*, number 7132 in Springer Lecture Notes in AI, pages 32–49, Berlin, 2012. Springer Verlag.

[74] J. P. Wallner. Structural constraints for dynamic operators in abstract argumentation. *Argument and Computation*, 11:151–190, 2020.

[75] Y. Wu and M. Caminada. A labelling-based justification status of arguments. *Studies in Logic*, 3:12–29, 2010.

[76] Y. Wu and M. Podlaszewski. Implementing crash-resistence and non-interference in logic-based argumentation. *Journal of Logic and Computation*, 25:303–333, 2015.

[77] 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.