On the Graded Acceptability of Arguments in Abstract and Instantiated Argumentation[☆]

Davide Grossi

University of Groningen

Sanjay Modgil

King's College London

Abstract

The paper develops a formal theory of the degree of justification of arguments, which relies solely on the structure of an argumentation framework, and which can be successfully interfaced with approaches to instantiated argumentation. The theory is developed in three steps. First, the paper introduces a graded generalization of the two key notions underpinning Dung's semantics: self-defense and conflict-freeness. This leads to a natural generalization of Dung's semantics, whereby standard extensions are weakened or strengthened depending on the level of self-defense and conflict-freeness they meet. The paper investigates the fixpoint theory of these semantics, establishing existence results for them. Second, the paper shows how graded semantics readily provide an approach to argument rankings, offering a novel contribution to the recently growing research programme on ranking-based semantics. Third, this novel approach to argument ranking is applied and studied in the context of instantiated argumentation frameworks, and in so doing is shown to account for a simple form of accrual of arguments within the Dung paradigm. Finally, the theory is compared in detail with existing approaches.

1. Introduction

Argumentation is a form of reasoning that makes explicit the reasons for the conclusions that are drawn, and how conflicts between reasons are resolved. Recent years have witnessed intensive study of both logic-based and human orientated models of argumentation and their use in formalising agent reasoning, decision making, and inter-agent dialogue [2, 3]. Much of this work builds on Dung's seminal theory of abstract

argumentation [4]. A Dung argumentation framework (AF) [4] is essentially a directed graph relating arguments by binary, directed forms of conflict, called attacks. The sceptically or credulously justified arguments are those in the intersection, respectively union, of sets—called extensions—of 'acceptable' arguments evaluated under various semantics (see [5] for an overview). Extensions are evaluated based on two core principles. Firstly, a set of arguments should not contain internal attacks, that is, it should be *conflict-free*. Secondly, it should defend itself in the sense that any argument a in the set is either un-attacked, or if attacked by some argument b, there is then an argument in the set that defends a by attacking b (in which case a is said to be defended by, or *acceptable with respect to*, the set of arguments). Arguments and attacks may be seen as primitive or assumed to be defined by an instantiating set of sentences in natural language or in a formal logical language. The former case thus provides for characterisations of more human-orientated uses of argument in reasoning, while in the latter case, the claims of the justified arguments identify the non-monotonic inferences from the instantiating set of logical formulae, thus providing for dialectical characterisations of non-monotonic logics [4].

Dung's theory has been extended in a number of directions. In particular, 'exogenously' given information about the relative strength of arguments has been used to determine which attacks succeed as defeats, so that the acceptable arguments are evaluated with respect to the arguments related by defeats rather than attacks. In this way one can effectively arbitrate amongst credulously justified conflicting arguments. Examples include [6, 7, 8] that make use of a preference relation over arguments (where in the case of [7], preferences are expressed by arguments that attack attacks amongst arguments) and [9] that makes use of an ordering over the values promoted by arguments. Other notable developments include approaches that associate probabilities with arguments [10, 11], and weights on attacks so that extensions do not necessarily comply with the conflict-freeness requirement on Dung extensions; arguments in an extension may attack each other provided that the summative weight of attacks does not exceed a given 'inconsistency budget' [12].[1]

Context: graduality in argumentation. It has long been recognized—since at least [14] and [15]—that one of the drawbacks of Dung's theory of abstract argumentation is the limited level of granularity the theory offers in differentiating the strength or status of arguments (essentially three, but cf. [16]). Consider the following informal example:

Example 1. *Suppose an argument I concluding the presumed innocence of a suspect. I is attacked by an argument G which consists of two sub-arguments G_1 and G_2 that respectively conclude that a suspect had* opportunity *and* motive*, where G defeasibly extends these sub-arguments to conclude that the suspect is guilty. Suppose an argument I_1 (in support of an alibi) that attacks G_1 on the assumption that the suspect does not have an alibi. Suppose then an additional argument I_2 that attacks an assumption in G_2. The level to which I is defended, and so said to be justified, is intuitively increased in the case that we have counter-arguments I_1 and I_2 that argue against*

---

[1]See also [13] which account for the relative strength of attacks amongst arguments.

the suspect having motive and *opportunity, as compared with when one only has the counter-argument I_1.*

The above example highlights one amongst a number of intuitions that are formalised by the above mentioned [14] and [15], and more recently [17, 18, 19], in which the numbers of attackers and defenders are used to give a more fine grained assignment of status to (and hence ranking of) arguments. While these approaches are defined wth respect to AFs, the status of arguments is not determined using the standard 'Dungian' concepts of sets of arguments (extensions) and arguments defended by these sets; rather, measures of the strengths of arguments based on the numbers of attackers and/or defenders are propagated through the AF graph[2]. These approaches, which we refer to here as 'propagation based approaches' (see [20] for a recent comparative overview), have also been developed to account for exogenously given information about the strength of arguments, which provide the initial measures that are then adjusted based on how they are propagated through the graph. Notable examples of the latter include [21, 22] and [23].

Paper Contribution. The central aim of this paper is to show that a more fine grained assignment of status to arguments that does not rely on exogenous information, and thus only on the numbers of attackers and defenders of arguments, can be formalised as a natural generalisation of the Dungian notions of conflict free sets of arguments and the defense of arguments by sets of arguments. This aim is achieved in the following way. *First*, our starting point is the classical definition of an admissible set of arguments as one that is conflict free and that defends all its contained arguments [4]. We show that the logical structure of this definition naturally generalises so as to yield more fine grained graded notions of conflict freeness and defense that account for the number of attackers and defenders, thereby obtaining a graded variant of the concept of admissibility. *Second*, this graded form of admissibility serves as a basis for defining graded variants of all the classic semantics of abstract argumentation studied in [4]: complete, grounded, stable and preferred. Dung's definitions of the standard semantics can all be retrieved as special cases of our graded variants, showing that the form of graduality this paper studies is rooted in a principled way in the classical theory of abstract argumentation. These graded semantics are, intuitively, ways of interpreting the standard Dung semantics in 'stricter' or 'looser' ways. For instance, the grounded semantics can be interpreted more 'strictly' by requiring that all attackers be counterattacked by at least two arguments, instead of just one as in the classic case. So each Dung semantics now comes equipped with a family of strengthenings and weakenings, which we call *graded semantics*. *Third*, these strengthenings and weakenings define a natural (partial) ordering dictated by set-inclusion: a stricter semantics will define sets of arguments which are subsets of the sets defined by a weaker one. This natural ordering then induces an ordering on the arguments themselves, thereby defining a ranking over the arguments in the framework. Such a ranking enables arbitration amongst arguments without recourse to exogenous preference information. As this

---

[2]The exception being [19] in which the strength of arguments is evaluated by reference to two person games in which a proponent defends an argument against counter-attacks by an opponent.

ranking is induced from a generalization of Dung standard semantics, our approach is at the outset methodologically distinct from propagation-based approaches, where argument rankings are defined directly by reference to the argument graph. *Fourth*, we study the application of graded semantics and their induced rankings to two key instantiations of Dung AFs: classical logic instantiations that, under the standard Dung semantics, yield a dialectical characterisation of non-monotonic inference in Preferred Subtheories [24], and instantiations that accommodate more human orientated uses of argumentation through the use of schemes and critical questions [25]. In so doing, we seek to substantiate some of the intuitions captured by our generalisation of Dung semantics, and show how the graded semantics capture a simple form of counting based accrual of arguments, which has traditionally been regarded as being incompatible with Dung's theory [26].

Outline of the paper. The paper is structured in three parts. $\boxed{\text{Part 1}}$ concerns the development of the abstract theory of graded argumentation. It starts with Section 2, where we review Dung's theory, giving prominence to its fixpoint-theoretic underpinnings, which have remained relatively under-investigated in the literature, and that we thus consider to be of some independent interest. Section 3 then generalises Dung's notions of conflict freeness and defense, yielding grading variants of these notions in terms of the number of arguments attacking and defending any given a whose acceptability with respect to a given set of arguments is at issue. This yields a ranking among types of conflict freeness and defense. Section 4 then generalises Dung's standard semantics, so that extensions are graded with respect to the attacks and counter-attacks on their contained arguments. These semantics—which we call *graded*—are shown to generalise Dung's theory and are studied providing constructive existence results. $\boxed{\text{Part 2}}$ first shows, in Section 5, how the new graded semantics yield a natural way of ranking arguments according to how strongly they are justified under different graded semantics, thereby enabling endogenous arbitration among credulously justified arguments. Then, Section 6 illustrates application of these type of rankings to *ASPIC+* instantiations of AFs [8] that formalise stereotypical patterns of argumentation encoded in schemes and critical questions [25], thus accounting for more human-orientated uses of argument, and *ASPIC+* instantiations of AFs that provide dialectical characterisations of non-monotonic inference in Preferred Subtheories [24]. Both types of instantiation are then shown to capture a simple form of counting based accrual, whereby multiple arguments in support of the same conclusion mutually strengthen each other. In $\boxed{\text{Part 3}}$, Section 7 develops a thorough comparison of our approach to the existing approaches to graduality and rankings in argumentation, leveraging the systematization recently introduced in [20]. This allows us to place more precisely graded argumentation in the growing landscape of ranking-based semantics. We conclude in Section 8 outlining some avenues for future research in graded argumentation.

2. Preliminaries: Abstract Argumentation

This preliminary section reviews key concepts and results from Dung's abstract argumentation theory [4]. The presentation we provide gives prominence to the fixpoint
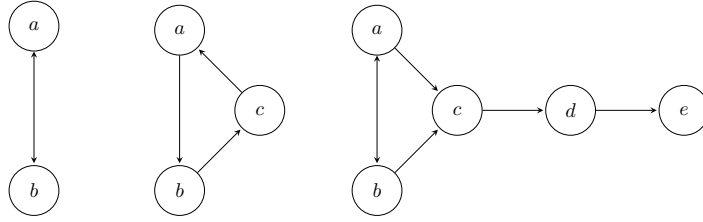
4

Figure 1: Argumentation frameworks

theory underpinning Dung's theoretical framework. After [4] the fixpoint theory of abstract argumentation has remained relatively under-investigated in the literature. It is, however, the most natural angle from which to pursue the objectives of this paper. We are unaware of any comprehensive exposition to date of the fixpoint theory of abstract argumentation, and so hope this preliminary section is of independent interest.

Since this paper will provide a generalization of Dung's original theory, all results presented in this section can actually be obtained as direct corollaries of the results we will establish later in Section 4. This, we argue, should be a desirable feature for any theory of graduality in argumentation which bases itself on Dung's original proposal. For completeness of the exposition, direct proofs of the results dealt with in this section can be found in Section 8.

2.1. Basic Definitions

Definition 1 (Frameworks)**.** *An* argumentation framework *(AF) Δ is a tuple $\langle \mathcal{A}, \rightarrow \rangle$ where $\mathcal{A} \neq \emptyset$, and $\rightarrow \, \subseteq \mathcal{A}^2$ is a binary attack relation on $\mathcal{A}$. Notation $x \rightarrow y$ denotes that x attacks y, and $X \rightarrow x$ denotes that $\exists y \in X$ s.t. $y \rightarrow x$. Similarly, $x \rightarrow X$ denotes that $\exists y \in X$ s.t. $x \rightarrow y$. For a given Δ we write $\overline{x}$ to denote $\{y \in \Delta \mid y \rightarrow x\}$ (the direct attackers of x), and $\overline{\overline{x}}$ to denote $\{z \in \Delta \mid z \rightarrow y, y \in \overline{x}\}$ (the direct defenders of x). An AF such that for each x, $\overline{x}$ is finite, is called* finitary*, whereas an AF with a finite number of arguments $\mathcal{A}$ is called* finite*, else* infinite*.*

We will sometimes refer to AFs as *attack graphs*.[3] Figure 1 depicts three AFs. An argument $a \in \mathcal{A}$ is said to be acceptable w.r.t. $X \subseteq \mathcal{A}$, if any argument attacking a is attacked by some argument in X, in which case X is said to defend a. An AF's characteristic (also called 'defense') function, applied to some $X \subseteq \mathcal{A}$, returns the arguments defended by X [4]:

Definition 2 (Defense Function)**.** *The defense function $\mathsf{d}_\Delta : \wp(\mathcal{A}) \longrightarrow \wp(\mathcal{A})$ for $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ is defined as follows. For any $X \subseteq \mathcal{A}$:*

$$\mathsf{d}_\Delta(X) \quad = \quad \{x \in \mathcal{A} \mid \forall y \in \mathcal{A} : \text{IF } y \rightarrow x \text{ THEN } X \rightarrow y\}$$

Where no confusion arises we may drop the subscript Δ in d_Δ.

---

[3]Although note that all definitions in this paper equally apply to 'defeat' graphs which assume a binary defeat relation on arguments, obtained through use of preferences in deciding which attacks succeed as defeats.

An argument $a \in \mathcal{A}$ is not attacked by a set $X \subseteq \mathcal{A}$ if no argument in X attacks a. One can define a function which, applied to some $X \subseteq \mathcal{A}$ in an AF, returns the arguments that are not attacked by X. This function was introduced by Pollock in [27] for his theory of defeasible reasoning, and we refer to it here as the 'neutrality function'.

Definition 3 (Neutrality Function). *The neutrality function* $\mathrm{n}_\Delta : \wp(\mathcal{A}) \longrightarrow \wp(\mathcal{A})$ *for* $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ *is defined as follows. For any* $X \subseteq \mathcal{A}$:

$$\mathrm{n}_\Delta(X) = \{ x \in \mathcal{A} \mid \text{ NOT } X \rightarrow x \}.$$

Again, where no confusion arises we may drop the subscript Δ *in* n_Δ.

One final bit of terminology. In what follows we will often use the notion of function iteration for d and n which we define in the standard inductive way, for $F \in \{\mathrm{d}, \mathrm{n}\}$: $F^0(X) = X$; $F^{k+1}(X) = F(F^k(X))$.

Example 2 (Defense and neutrality in Figure 1). *The functions applied to the symmetric graph of Figure 1 (left) yield the following equations:*

$$
\begin{array}{rclcrcl}
\mathrm{d}(\emptyset) & = & \emptyset & \quad & \mathrm{n}(\emptyset) & = & \{a, b\} \\
\mathrm{d}(\{a\}) & = & \{a\} & & \mathrm{n}(\{a\}) & = & \{a\} \\
\mathrm{d}(\{b\}) & = & \{b\} & & \mathrm{n}(\{b\}) & = & \{b\} \\
\mathrm{d}(\{a, b\}) & = & \{a, b\} & & \mathrm{n}(\{a, b\}) & = & \emptyset
\end{array}
$$

Notice that the output of n *on* $\emptyset$ *corresponds to the whole set of arguments, as no arguments can be attacked by* $\emptyset$. *Notice also that while* $\{a\}$ *and* $\{b\}$ *are included in* $\mathrm{n}(\{a\})$ *and* $\mathrm{n}(\{b\})$, $\{a, b\}$ *is not.*

One can define the extensions of an AF Δ under Dung's semantics, in terms of the fixpoints ($X = \mathrm{d}_\Delta(X)$ and $X = \mathrm{n}_\Delta(X)$) or post-fixpoints ($X \subseteq \mathrm{d}_\Delta(X)$ and $X \subseteq \mathrm{n}_\Delta(X)$) of the defense and neutrality functions, as recapitulated in Table 1. The justified arguments are then defined under various semantics:

Definition 4 (Justification under Semantics). *Let* $\Delta = \langle \mathcal{A}, \rightarrow \rangle$. *Then for semantics* $S \in \{grounded, stable, preferred\}$[4], $a \in \mathcal{A}$ *is* credulously, *respectively* sceptically, justified *under* S, *if* a *is in at least one, respectively all,* S *extensions of* Δ.

Finally, we recapitulate some well-known properties, first established in [4], of the defense and neutrality functions that will be referred to later:

Fact 1. *Let* $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ *be an* AF *and* $X, Y \subseteq \mathcal{A}$. *The following holds:*

$$
\begin{array}{rcl}
X \subseteq Y & \implies & \mathrm{d}(X) \subseteq \mathrm{d}(Y) \\
X \subseteq Y & \implies & \mathrm{n}(Y) \subseteq \mathrm{n}(X) \\
\mathrm{d}(X) & = & \mathrm{n}(\mathrm{n}(X))
\end{array}
$$

---

[4]Typically, the justified arguments are not defined w.r.t. the complete semantics, which subsume each of grounded, stable and preferred.

X is conflict-free in Δ	iff	$X \subseteq \mathtt{n}_\Delta(X)$
X is self-defended in Δ	iff	$X \subseteq \mathtt{d}_\Delta(X)$
X is admissible in Δ	iff	$X \subseteq \mathtt{n}_\Delta(X)$ and $X \subseteq \mathtt{d}_\Delta(X)$
X is a complete extension in Δ	iff	$X \subseteq \mathtt{n}_\Delta(X)$ and $X = \mathtt{d}_\Delta(X)$
X is a stable extension of Δ	iff	$X = \mathtt{n}_\Delta(X)$
X is the grounded set in Δ	iff	X is the smallest complete ext. of Δ ($X = \mathsf{lfp}.\mathtt{d}_\Delta$)
X is a preferred extension of Δ	iff	X is a largest complete extension of Δ

Table 1: Classical notions of abstract argumentation theory from [4].

That is, function $\mathtt{d}$ is monotonic, function $\mathtt{n}$ is antitonic, and the composition of $\mathtt{n}$ with itself, which we will also denote $\mathtt{n} \circ \mathtt{n}$, is function $\mathtt{d}$. For example, in Figure 1 (right), we have that $\mathtt{n}(\{a\}) = \{a, d, e\}$, and $\mathtt{n}(\{a, d, e\}) = \{a, d\} = \mathtt{d}(\{a\})$.

Fact 2 (ω-continuity[5])**.** *If Δ is finitary, then $\mathtt{d}_\Delta$ is (upward-)continuous for any $X \subseteq \mathcal{A}$, i.e., for any upward directed set $D \in \wp(\wp(\mathcal{A}))$ of finite subsets of $\mathcal{A}$:*[6]

$$\mathtt{d}_\Delta \left(\bigcup_{X \in D} X \right) = \bigcup_{X \in D} \mathtt{d}_\Delta(X). \tag{1}$$

Similarly, $\mathtt{d}_\Delta$ is (downward-)continuous for any $X \subseteq \mathcal{A}$, i.e., for any downward directed set $D \in \wp(\wp(\mathcal{A}))$:

$$\mathtt{d}_\Delta \left(\bigcap_{X \in D} X \right) = \bigcap_{X \in D} \mathtt{d}_\Delta(X). \tag{2}$$

The above fact establishes important properties of the behaviour of the defense function with respect to sequences of sets of arguments and their limits. As we will also later see in the graded generalization of Dung's theory, these properties are key in the construction of extensions through the iteration of the defense function.

2.2. The Fixpoint Theory of Acceptability and Conflict-freeness

We show how any admissible set of arguments can be saturated to a complete extension through a process of fixpoint approximation. This establishes a general result concerning the computation of complete extensions in (finitary) attack graphs which, to the best of our knowledge, has never been reported in the literature. It is, however, a generalization of well-known existing results such as [4, Lemma 46, Theorem 47] (cf. also [28]).

---

[5]Cf. [4, Lemma 28].

[6]We recall that an upward directed set $D \in \wp(\wp(\mathcal{A}))$ is a set of sets such that any two elements X and Y in D have an upper bound in D, that is, there also exists a superset $Z \supseteq X \cup Y$ in D. A downward directed set is defined dually in the obvious way.

2.2.1. Construction of Fixpoints from Admissible Sets

Fix a framework Δ and take a set X such that $X \subseteq \mathtt{d}(X)$ and $X \subseteq \mathtt{n}(X)$ (i.e., an admissible set). By iterating $\mathtt{d}$, consider the stream of sets $\mathtt{d}^0(X), \mathtt{d}^1(X), \ldots$, and $\mathtt{d}^0(\mathtt{n}(X)), \mathtt{d}^1(\mathtt{n}(X)), \ldots$. Since X is admissible, $\mathtt{d}$ is monotonic and $\mathtt{n}$ antitonic (Fact 1), the first stream (see the lower stream in Figure 2) is non-decreasing and the second stream (see the upper stream in Figure 2) is non-increasing, with respect to set inclusion. In finite attack graphs, these streams must therefore stabilize reaching a limit at state $|\mathcal{A}| + 1$. In infinite but finitary attack graphs, we will see that the limit can be reached at ω. We will see (Lemma 2) that the limits of these streams correspond to the *smallest* fixpoint of $\mathtt{d}$ *containing* the admissible set X and, respectively, the *largest* fixpoint of $\mathtt{d}$ which is *contained* in $\mathtt{n}(X)$. We denote the first one by $\mathsf{lfp}_X.\mathtt{d}$ and the second one by $\mathsf{gfp}_X.\mathtt{d}$. Intuitively, the two sets denote the smallest superset of X which is equal to the set of arguments it defends[7] and, respectively, the largest set which is not attacked by X and which is equal to the set of arguments it defends.[8] The construction is illustrated in Figure 2 below.

Example 3. *Consider the cycle of length three in Figure 1 (center), and take the admissible set $\emptyset$. By applying the above construction we obtain immediately $\mathtt{d}(\emptyset) = \emptyset$ as limit of the lower stream, and $\mathtt{n}(\emptyset) = \{a, b, c\} = \mathtt{d}(\{a, b, c\})$ as limit of the upper stream. $\emptyset$ is the smallest self-defended set containing $\emptyset$, $\{a, b, c\}$ is the largest self-defended set contained in $\mathtt{n}(\emptyset)$.*

We prove now the correctness of the above construction by showing that the limits of the above streams correspond indeed to the desired fixpoints. First of all the following important lemma shows how conflict-freeness is preserved by the above process of iteration of the defense function:

Lemma 1. *Let Δ be a finitary attack graph and $X \subseteq \mathcal{A}$ be admissible. Then for any n s.t. $0 \le n < \omega$,*
$$X \subseteq \mathtt{d}_\Delta^n(X) \subseteq \mathtt{n}_\Delta(\mathtt{d}_\Delta^n(X)) \subseteq \mathtt{n}_\Delta(X).$$

That is, each $\mathtt{d}_\Delta^n(X)$ in the stream of iteration of the defense function from an admissible set X is a conflict-free set. The lemma can be seen as a reformulation of [4, Lemma 10], known as Dung's *fundamental lemma*.[9]

We can show that the above streams obtained through the process of iteration of the defense function construct the desired fixpoints:

---

[7]Theorem 1 will show this set is also conflict-free and it is therefore the smallest complete extension containing X.

[8]Note that such a set is not necessarily conflict-free. E.g., consider $\Delta = \langle \{a, b, c\}, \{(b, c), (c, b)\} \rangle$, that is $b \to c$ and $c \to b$. Then $\mathtt{n}(\{a\}) = \{a, b, c\}$ and $\mathtt{d}(\{a, b, c\}) = \{a, b, c\}$. But clearly it is not the case that $\{a, b, c\} \subseteq \mathtt{n}\{a, b, c\}$, that is, $\{a, b, c\}$ is not conflict-free.

[9]It also generalises [4, Lemma 46] to the case of X admissible, instead of $X = \emptyset$.
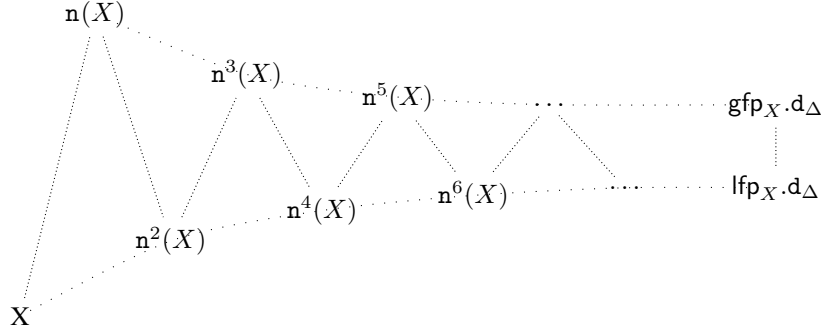
Figure 2: Streams generated by indefinite iteration of $\mathtt{n}$ applied to an admissible set X (in symbols, $\mathtt{n}^\omega(X)$). Position with respect to the horizontal axis indicates the number of iterations (growing from left to right), while position with respect to the vertical axis indicates set theoretic inclusion. The lower stream $X, \mathtt{n}^2(X), \mathtt{n}^4(X), \ldots$ stabilizes to $\mathsf{lfp}_X.\mathtt{d}_\Delta$. The upper stream $\mathtt{n}(X), \mathtt{n}^3(X), \mathtt{n}^5(X), \ldots$ stabilizes to $\mathsf{gfp}_X.\mathtt{d}_\Delta$.

Lemma 2. *Let Δ be a finitary attack graph and $X \subseteq \mathcal{A}$ be admissible:*

$$\mathsf{lfp}_X.\mathtt{d}_\Delta = \bigcup_{0 \leq n < \omega} \mathtt{d}_\Delta^n(X) \tag{3}$$

$$\mathsf{gfp}_X.\mathtt{d}_\Delta = \bigcap_{0 \leq n < \omega} \mathtt{d}_\Delta^n(\mathtt{n}_\Delta(X)) \tag{4}$$

Notice that since $\mathtt{n}^2 = \mathtt{d}$ (Fact 1), a stream generated by the indefinite iteration of the defense function can actually be viewed as a stream generated by the indefinite iteration of the neutrality function. So equations (3) and (4) of Theorem 2 can be rewritten as follows:

$$\mathsf{lfp}_X.\mathtt{d}_\Delta = \bigcup_{0 \leq n < \omega} (\mathtt{n}_\Delta^2)^n(X) \tag{5}$$

$$\mathsf{gfp}_X.\mathtt{d}_\Delta = \bigcap_{0 \leq n < \omega} (\mathtt{n}_\Delta^2)^n(\mathtt{n}_\Delta(X)) \tag{6}$$

In this light, Lemmas 1 and 2 capture several of the key features of the stream generated by the indefinite iteration of the neutrality function on an admissible set X. First, the stream can be split into two parts, the part consisting of even and, respectively, odd iterations of $\mathtt{n}_\Delta$. Second, the stream of even iterations converges to a limit which is the smallest complete set including X, and the stream of odd iteration converges to a limit which is the largest self-defended set contained in $\mathtt{n}_\Delta(X)$ (that is, not attacked by X).[10] Notice that such a set is just free of conflict with respect to X, but it is not necessarily conflict-free, and hence it is not necessarily a complete extension. Third,

---

[10]Again the finitariness assumption in the theorem could be lifted by making use of transfinite induction. Cf. Remark 1 below.

the two streams can actually be viewed as streams of the defense function $\mathtt{d}_\Delta$ applied to X and, respectively, to $\mathtt{n}_\Delta(X)$. Fourth, the two parts grow towards each other as the stream of even iterations is increasing, while the one of odd iterations is decreasing. See Figure 2 for an illustration.

Remark 1. *The proof of Lemma 2 relies in an essential manner on the finitariness assumption on the underlying framework. The assumption simplifies the proof but, it should be stressed, could be lifted. For infinite graphs which are not finitary, the lemma could be proved by resorting to transfinite induction:*

$$
\begin{aligned}
\mathtt{d}_\Delta^0(\emptyset) &= X \\
\mathtt{d}_\Delta^{\alpha+1}(\emptyset) &= \mathtt{d}_\Delta(\mathtt{d}_\Delta^\alpha(X)) \\
\mathtt{d}_\Delta^\lambda &= \bigcup_{\alpha<\lambda} \mathtt{d}_\Delta^\alpha(X) \quad \textit{(for λ arbitrary limit ordinal).}
\end{aligned}
$$

By the monotonicity of $\mathtt{d}_\Delta$ it can then be shown that there exists an ordinal α of cardinality at most $|A|$ such that: $\mathsf{lfp}_X.\mathtt{d}_\Delta = \mathtt{d}_\Delta^\alpha(X)$. A proof of this statement in the general setting of complete partial orders can be found in [29, Ch. 3]. Transfinite induction relies on the Axiom of Choice (or equivalent formulations such as Zorn's Lemma or the Well-Ordering Principle), which is known to be required for the existence results of the standard Dung semantics (cf. [30]).

2.2.2. Construction of Complete Extensions

With the above results in place, one can then show how complete extensions can be constructed through a process of fixpoint approximation.

Theorem 1. *Let Δ be a finitary AF and $X \subseteq \mathcal{A}$ be admissible. Then the limit $\bigcup_{0 \le n < \omega} \mathtt{d}_\Delta^n(X)$ is the smallest complete extension of Δ that includes X.*

The theorem establishes that, in finitary frameworks, any complete set can be computed via a process of iteration at ω of the defense function, starting with some admissible set. It is a novel simple generalization of the earlier result in [4] for the case of $X = \emptyset$. This process starts by including the arguments that have no attackers or that belong to an initial admissible set, then including those arguments that are defended by the first set of arguments included, and so on.[11] At an intuitive level, the theorem states that the indefinite iteration of the defense from an admissible X can be considered as a formalization of the process whereby an agent constructs a rational argumentative position—a complete extension—starting from X.

2.2.3. Construction of Other Dung Extensions

Theorem 1 also yields a constructive proof of existence (in finitary graphs) for the grounded extension. By setting $X = \emptyset$ (the trivially admissible set), the theorem returns the known result for the construction of the grounded extension $\mathsf{lfp}.\mathtt{d}_\Delta =$

---

[11]Cf. Remark 1.

$\bigcup_{0 \leq n < \omega} \mathtt{d}_\Delta^n(\emptyset)$ [4, Th. 47]. In this section we show how the theorem relates to the other classical Dung extensions.

Specific conditions can be identified which guarantee that the indefinite iteration of the defense function constructs preferred and stable extensions from a given admissible set X. It is fairly easy to see that if the chosen admissible set X of Δ is 'big enough' in the precise sense that it contains enough arguments to be able, from some argument in X, to reach any argument in the graph via the attack relation, i.e., if $\mathcal{A} \subseteq \{a \mid \exists b \in X : a \leftarrow^+ b\}$,[12] then the stream of iterations of $\mathtt{d}_\Delta$ from X converges to a complete extension containing X (by Theorem 1), but this extension is now maximal as all arguments in Δ can be reached from X.

The condition under which Theorem 1 constructs stable extensions is particularly interesting. If the streams of even and odd iterations of the neutrality function (recall Figure 2) converge to the same limit, then the process of fixpoint approximation defines a stable extension:

Fact 3. *Let Δ be a finitary AF and $X \subseteq \mathcal{A}$ be admissible. If $\bigcup_{0 \leq n < \omega} \mathtt{d}_\Delta^n(X) = \bigcap_{0 \leq n < \omega} \mathtt{d}_\Delta^n(\mathtt{n}_\Delta(X))$, then both sets coincide with the unique stable extension of Δ that includes X.*

Proof. By Theorem 1, $\bigcup_{0 \leq n < \omega} \mathtt{d}_\Delta^n(X)$ is the smallest complete extension containing X. However, as $\bigcup_{0 \leq n < \omega} \mathtt{d}_\Delta^n(X) = \bigcap_{0 \leq n < \omega} \mathtt{d}_\Delta^n(\mathtt{n}_\Delta(X))$ by assumption, the set $\bigcup_{0 \leq n < \omega} \mathtt{d}_\Delta^n(X)$ is therefore also a fixpoint of the neutrality function, and therefore a stable extension. $\square$

This observation is, to the best of our knowledge, novel and provides a characterization of the existence of a stable extension that includes a given admissible set.

Example 4 (Construction of complete extensions in Figure 1)**.** *Consider the rightmost AF. Starting with the admissible set $\{a\}$, the non-decreasing stream*

$$\{a\}, \{a, d\}, \{a, d\}, \dots$$

converges after one step to the smallest complete extension $\mathsf{lfp}_{\{a\}}.\mathtt{d} = \{a, d\}$ containing $\{a\}$. The non-increasing stream $\{a, d, e\}, \{a, d\}, \{a, d\}, \dots$ converges to $\mathsf{gfp}_{\{a\}}.\mathtt{d} = \{a, d\}$, i.e., to the same set which is also the largest fixpoint of $\mathtt{d}$ included in $\mathtt{n}(\{a\}) = \{a, d, e\}$. As $\{a, d\}$ is also conflict-free, it is a preferred extension. Notice also that if we were to start with $\emptyset$, the resulting streams would be $\emptyset, \emptyset, \dots$ and $\{a, b, c, d, e\}$, $\{a, b, c, d, e\}, \dots$. The first one constructs the grounded extension $\emptyset$ of the AF, and the second the largest fixpoint of $\mathtt{d}$ in the AF, that is, $\{a, b, c, d, e\}$.

3. Graded Acceptability

We now turn to the main contribution of this paper: a graded generalisation of Dung's acceptability semantics. We first introduce the intuitions behind our generalisation, and then define and study graded variants of Dung's defense and neutrality

---

[12]As customary, $+$ denotes the transitive closure of a given relation, in our case the attack relation $\leftarrow$. So $a \leftarrow^+ b$ stands for 'there exists a path of attacks from b to a'.
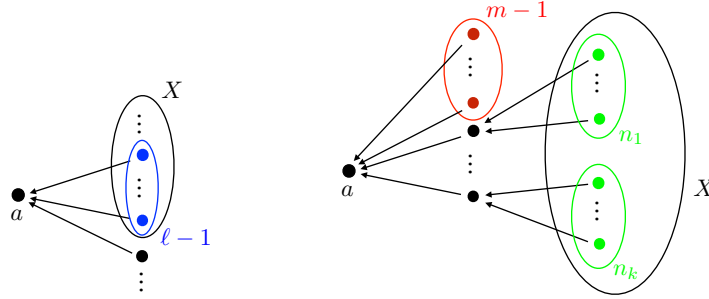
Figure 3: Variants of neutrality (left) and defense (right). The lower integer ℓ is, the more neutral is the set of arguments X with respect to a (left). The lower integer m is and the higher integer $n = \min(n_1, \ldots, n_k)$ is, the better is a defended by X (right). Standard neutrality and defense correspond to $\ell = m = n = 1$.

functions, which capture the proposed intuitions. These functions will then be used later (Section 4) to define and study a family of graded variants of Dung's semantics, and the rankings they enable (Section 5).

3.1. Introducing Graded Acceptability: Intuitions

The central tenet of argumentation theory is that any individual argument cannot, in and of itself, constitute definitive grounds for believing that a claim is true. Rather, the status of an epistemic claim as true justified belief is not established by an individual argument, but through the dialectical consideration of counter-arguments and defenders of these counter-arguments [31]. Similarly in practical reasoning, the status of (a claim representing) a decision option supported by an individual practical argument is not considered to be the option that simply maximises a given objective, but rather the best option contingent on having dis-preferred alternative options and refuted challenges made to the assumptions made in support of the argument. Pragmatically however, we operate under the assumption that the claim is true/the decision option is the best available, to the extent that as of yet we know of no good reason to suppose otherwise. Argumentatively, a claim can be considered established only in as much as it is the claim of a justified argument included in a network of interrelated arguments and counter-arguments. Dung's abstract argumentation theory captures these principles by assuming *sets of* arguments, rather than individual arguments, as the units of analysis, and studying formal criteria (semantics) for sets of arguments to be acceptable. Apart from trivial cases (unattacked arguments), arguments are acceptable only as members of a set of acceptable arguments. The graded theory of acceptability that we aim at, captures a notion of graduality while at the same time retaining the notion of a set of arguments as the central unit of analysis.

3.1.1. Graded neutrality

According to the standard definition of neutrality (Definition 3) a set X is neutral with respect to a if a is not attacked by *any* argument in X. A less demanding criterion of neutrality of X with respect to a would require that there exists *at most one* attacker

12

of a in X, a yet less demanding one (or at least not 'as demanding as') would require that there exist *at most two* attackers of a in X, and so on. Intuitively, these weakened neutrality criteria capture the idea that one (two, three, . . .) attackers are not enough to rule out the co-acceptability of a and its attacking arguments.[13] One then obtains a natural way to generalise the neutrality function (Definition 6 below) by making explicit a numerical level ℓ of neutrality of a set of arguments with respect to a given argument, as depicted in Figure 3 (left).[14] So we say that a is ℓ-neutral with respect to X whenever there are at most $\ell - 1$ attackers of a in X.

3.1.2. Graded defense

According to the standard notion of defense, an argument a is defended by a set of arguments X whenever *every* attacker of a is attacked by *some* argument in X. The quantification pattern ('for all', 'some') involved in this definition offers again a natural handle to generalise the notion of defense. If *all but at most one* attackers of a are attacked by at least one argument in X, the quality of such defense (and hence the extent to which a is acceptable w.r.t. X) can reasonably be considered 'lower' than in the case in which *all* arguments are counter-attacked by at least one argument in X. But the former quality of defense is still 'higher' than (or at least not 'as low as') the case in which *all but at most two* attackers are counter-attacked by at least one argument in X, and so on. Similarly, if all attackers of a are counterattacked by *at least two* arguments in X, then the quality of this defense can reasonably be considered 'higher' than in standard acceptability, but 'lower' than (or at least not 'as high as') the case in which all attackers of a are counterattacked by *at least three* arguments in X. Combining these intuitions—depicted in Figure 3 (right)—one obtains a way to generalise the defense function (Definition 5 below) by making explicit, through numeric grades (m and n) of the above type, how well a set X defends an argument a. So we say that X mn-defends a whenever there are at most $m - 1$ attackers of a, which are not counterattacked by at least n arguments in X.

This notion of graded defense is related in a natural way to the above notion of graded neutrality: the set of arguments that are mn-defended by X, is the set of arguments which is not attacked by at least m arguments, that are not in turn attacked by at least n arguments in X (i.e., arguments that are m-neutral with respect to the set of arguments that are n-neutral with respect to X). In other words, the notion of tolerance towards attack (graded neutrality) can be iterated to obtain a notion of graded defense. Fact 4 will establish this claim formally. We illustrate the above intuitions with a few examples.

---

[13]Using terminology from logic, this may be viewed as an argumentative form of paraconsistency. In more 'human orientated' argumentation formalisms (e.g., [12] and developments thereof), this may be viewed as accounting for an attacking argument not establishing definitive grounds for its claim (as discussed at the beginning of Section 3.1), and hence not definitively ruling out the claim of the attacked argument.

[14]Weighted Argument Systems [12] propose a somewhat similar idea, whereby an inconsistency budget sets a threshold on the number of attacks that can be tolerated within a given set. However, notice that our notion of a threshold set, yielded by graded neutrality, is local in the sense that it pertains to the incoming attacks *on each individual argument*. We will later compare this and other related approaches in more detail (Section 7).

Example 5. *In Figures 4i) – 4iv), the encircled set Xi defends ai ($i = 1 \ldots 4$) under Dung's Definition 2. However we can differentiate these cases based on the number of attackers and defenders of ai. For instance, $X2$ more strongly defends $a2$ than $X1$ defends $a1$, as $a2$ is defended by two arguments whereas $a1$ is defended by one argument*

365 *(i.e., the standard of defense that allows at most 0 attackers to not be defended by 2 arguments is met by $X2$'s defense of $a2$ but not by $X1$'s defense of $a1$). We will later, in Example 12, reference the defense of $a3$ by $X3$ and of $a4$ by $X4$ to illustrate that neither can be said to be a more strong defense than the other. While neither $X5$ or $X6$ defend $a5$, respectively $a6$, under Dung's Definition 2, observe that $X5$'s defense*

370 *of $a5$ is stronger than $X6$'s defense of $a6$. The former meets a standard of defense that allows at most one attacker ($d5$) not to be defended by at least one defender (which goes hand in hand with accommodating the co-acceptability of $a5$ with at most one undefended attacker; i.e., $d5$ is 2-neutral with respect to $X5$). This standard is not met by $X6$'s defense of $a6$, since $a6$ is attacked by two undefended attacks (from $d6$ and*

375 *$e6$). In the latter case, a weaker standard of defense is met, which again goes hand in hand with accommodating the co-acceptability of $a6$ with its two undefended attackers. These notions then naturally generalise so that one can discriminate standards of defense based only on the number of attackers. Allowing at most one attacker not to be defended by two arguments, is a standard of defense met by $X1$s defense of $a1$, but not*

380 *$X3$'s defense of $a3$ (the former defense thus being stronger than the latter).*

3.2. Graded Defense and Neutrality Functions

Let us move now to the formal definitions of graded defense and neutrality. Take an argument a and a set of arguments X. Let k be the number of a's attackers $(b_1, \ldots, b_k)$ and, for each b_i ($1 \leq i \leq k$) let n_i be the (non zero) number of attackers

385 of b_i (i.e., defenders of a) in X. Finally, let n be the minimum among the n_is, i.e., $n = \min(\{n_i\}_{1 < i \leq k})$. We can now count the number m ($\leq k$) of attackers of a, which are counter-attacked by at least n arguments in X. Integers m and n therefore encode information about how strongly a is defended by X, in the sense that they express a maximum number (i.e., $m - 1$) of attackers of a which are not counterattacked by a

390 minimum given number (i.e., n) of arguments in X. We can now generalise Definition 2 as follows:

Definition 5 (Graded defense). *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF and let m and n be two positive integers ($m, n > 0$). The graded defense function $\mathrm{d}\underset{n}{m} : \wp(\mathcal{A}) \longrightarrow \wp(\mathcal{A})$ for Δ is defined as follows. For any $X \subseteq \mathcal{A}$:*

$$\mathrm{d}\underset{n}{m}(X) = \{x \in \mathcal{A} \mid \nexists_{\geq m} y : [\, y \rightarrow x \text{ AND } \nexists_{\geq n} z : [\, z \rightarrow y \text{ AND } z \in X \,] \,]\}$$

where $\exists_{\geq n} x$, for integers n ('there exist at least n arguments x') are the standard first-order logic counting quantifiers.[15] In the rare cases in which we need to make Δ

---

[15]Cf. [32]. The definition can be reformulated without counting quantifiers as follows:

$$\mathrm{d}\underset{n}{m}(X) = \{x \in \mathcal{A} \text{ S.T. } \mid \{y \in \overline{x} \text{ S.T. } \mid \{z \in \overline{y} \cap X\} \mid < n\} \mid < m\}$$

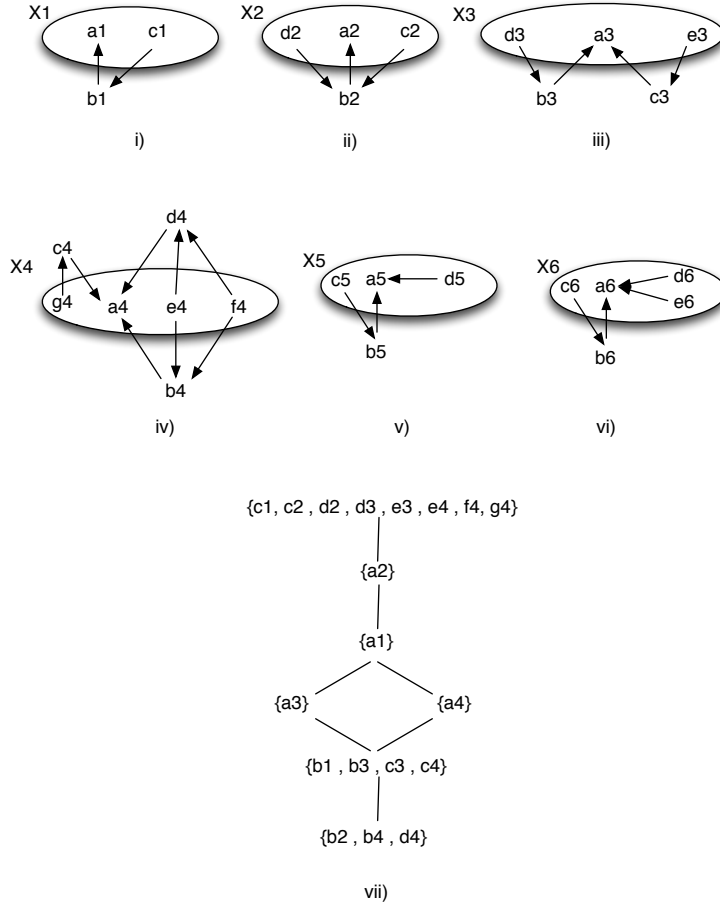where we write $\overline{x}$ to denote $\{y \in \mathcal{A} \mid y \rightarrow x\}$.

Figure 4: i)–vi) $X_{i \in \{1...6\}}$ defending $a_1, \ldots a_6$. vii) is the Hasse diagram of a preorder on the set of all arguments from i)–v) relevant for Example 12.

explicit we write $\text{d}\frac{\Delta}{m}_n$.

So, $\text{d}\frac{m}{n}(X)$ is the set of arguments (in the given framework) which have at most $m-1$ attackers that are not counter-attacked by at least n arguments in X.

Example 6. *In Figure 4,* $a1 \in \text{d}_{\frac{1}{1}}(X1)$ *and* $a2 \in \text{d}_{\frac{1}{1}}(X2)$ *since in both cases the following holds: at most 0 arguments attacking* $a1$*, respectively* $a2$*, are not attacked by at least one argument in* $X1$*, respectively* $X2$*. However if we increment n by 1 we have that:* $a2 \in \text{d}_{\frac{1}{2}}(X2)$ *but* $a1 \notin \text{d}_{\frac{1}{2}}(X1)$ *since it is the case that at least one argument attacking* $a1$ *is not attacked by at least two arguments in* $X1$*. Intuitively, this standard of defense allows for up to 0 attackers to not be counter-attacked by two defenders, a standard met by* $X2$*'s defense of* $a2$*, but not by* $X1$*'s defense of* $a1$*. Continuing with*

15

Figure 4, $a5 \in \mathtt{d}_{\underset{1}{3}}(X5)$ and $a6 \in \mathtt{d}_{\underset{1}{3}}(X6)$, since in both cases the standard of defense that allows for no more than 2 unattacked arguments is met. However, $a5 \in \mathtt{d}_{\underset{1}{2}}(X5)$ and $a6 \notin \mathtt{d}_{\underset{1}{2}}(X6)$ since this standard of defense accommodates up to a maximum of 1 unattacked attackers, and in the latter case there is more than one unattacked attacker of $a6$. Finally, $a1 \notin \mathtt{d}_{\underset{2}{1}}(X1)$ and $a3 \notin \mathtt{d}_{\underset{2}{1}}(X3)$ since the $\mathtt{d}_{\underset{2}{1}}$ standard of defense requires that all attackers of $a1$ ($a3$) are attacked by at least two arguments. However, $a1 \in \mathtt{d}_{\underset{2}{2}}(X1)$ and $a3 \notin \mathtt{d}_{\underset{2}{2}}(X3)$, since the standard of defense allowing at most one attacker not to be defended by two counter-attacks is met by $a1$ but not by $a3$. Notice that in this last case the two arguments $a1$ and $a3$ are discriminated based on the number of their attackers.

By the same logic, Definition 3 can be generalised as follows:

Definition 6 (Graded neutrality function). *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF and let ℓ be any positive integer. The graded neutrality function $\mathtt{n}_l : \wp(\mathcal{A}) \longrightarrow \wp(\mathcal{A})$ for Δ is defined as follows. For any $X \subseteq \mathcal{A}$:*

$$\mathtt{n}_\ell(X) = \{x \in \mathcal{A} \mid \nexists_{\geq \ell}\, y : y \rightarrow x \text{ AND } y \in X\}.$$

So, given a set of arguments X, $\mathtt{n}_\ell(X)$ denotes the set of arguments which have at most $\ell - 1$ attackers in X.[16]

Example 7. *In Figure 4, $n_2(X_5) = \{a5, b5, c5, d5\}$. Notice that $n_2(\{a5, b5, c5, d5\}) = \{b5, c5, d5\}$. Also, $n_2(X_6) = \{b6, c6, d6, e6\}$.*

3.3. Properties of Graded Defense and Neutrality

The following two facts show that the graded defense and neutrality functions are generalisations of the standard functions defined in Definitions 2 and 3, and that such generalisations remain well-behaved in the sense that they retain many of the key features of their standard variants.

Fact 4. *For any AF $\Delta = \langle \mathcal{A}, \rightarrow \rangle$, m, n and ℓ positive integers:*

$$\mathtt{d}_{\underset{1}{1}}(X) \quad = \quad \mathtt{d}(X) \tag{7}$$

$$\mathtt{n}_1(X) \quad = \quad \mathtt{n}(X) \tag{8}$$

$$X \subseteq Y \quad \Longrightarrow \quad \mathtt{n}_\ell(Y) \subseteq \mathtt{n}_\ell(X) \tag{9}$$

$$X \subseteq Y \quad \Longrightarrow \quad \mathtt{d}_{\underset{n}{m}}(X) \subseteq \mathtt{d}_{\underset{n}{m}}(Y) \tag{10}$$

$$\mathtt{n}_m(\mathtt{n}_n(X)) \quad = \quad \mathtt{d}_{\underset{n}{m}}(X) \tag{11}$$

---

[16]Equivalently, graded neutrality can be defined as follows, without the use of counting quantifiers:

$$\mathtt{n}_\ell(X) = \{x \in \mathcal{A} \text{ s.t. } |\bar{x} \cap X| < \ell\}.$$

Proof. Equation (7) follows from the fact that Definition 2 can be retrieved from Definition 5 by setting $n = m = 1$. Similarly (8) follows from the fact that Definition 3 can be retrieved from Definition 6 by setting $\ell = 1$. Equation (11) follows from Definitions 5 and 6 by the following series of equations:

$$
\begin{aligned}
\mathtt{n}_m(\mathtt{n}_n(X)) &= \mathtt{n}_m(\{y \in A \mid \nexists_{\geq n} z : [\, z \to y \text{ AND } z \in X]\}) \\
&= \{x \in A \mid \nexists_{\geq m} y : [\, y \to x \text{ AND } \nexists_{\geq n} z : [\, z \to y \text{ AND } z \in X]]\} \\
&= \mathtt{d}_n^m(X)
\end{aligned}
$$

Formulae (9) and (10) are direct consequences of Definitions 5 and 6. □

Equation (7) reformulates $\mathtt{d}(X)$ as the set of arguments for which it is not the case that there are one or more attackers, which are not counter-attacked by one or more arguments in X; that is, no attacker is not attacked by some argument in X. So does Equation (8) for $\mathtt{n}(X)$. The remaining formulae generalise Fact 1 to the graded setting. In particular, graded defense is monotonic (10), graded neutrality is antitonic (9), and equation (11) shows that, as in the standard case, the defense function is the two-fold iteration of the neutrality function (as in the standard case we may use the notation $\mathtt{n}_n \circ \mathtt{n}_m$ to denote this composition).

Importantly, the continuity of the defense function is also preserved in the graded setting:

Fact 5 (ω-continuity of graded defense). *If Δ is finitary, then function $\mathtt{d}_n^m$ is (upward-) continuous for any $X \subseteq A$, m and n positive integers. I.e., for any upward directed set D of finite subsets of $\mathcal{A}$:*

$$
\mathtt{d}_n^m\Big(\bigcup_{X \in \mathcal{D}} X\Big) = \bigcup_{X \in \mathcal{D}} \mathtt{d}_n^m(X) \tag{12}
$$

Similarly, $\mathtt{d}_n^m$ is (downward-)continuous for any $X \subseteq A$, m and n positive integers. I.e., for any downward directed set D of finite subsets of $\mathcal{A}$:

$$
\mathtt{d}_n^m\Big(\bigcap_{X \in \mathcal{D}} X\Big) = \bigcap_{X \in \mathcal{D}} \mathtt{d}_n^m(X) \tag{13}
$$

Sketch of proof. The argument used to prove Fact 2 carries through in exactly the same manner, exploiting the monotonicity of $\mathtt{d}_n^m$ (10) and the finitariness assumption over Δ. □

Finally, we establish some properties showing how the values for the defence and neutrality functions are affected by varying the parameters m and n.

Fact 6. *For any AF $\Delta = \langle \mathcal{A}, \to \rangle$, $X \subseteq \mathcal{A}$, and ℓ, m and n positive integers:*

$$
\begin{aligned}
\mathtt{n}_\ell(X) &\subseteq \mathtt{n}_{\ell+1}(X) & (14) \\
\mathtt{d}_n^m(X) &\subseteq \mathtt{d}_n^{m+1}(X) & (15) \\
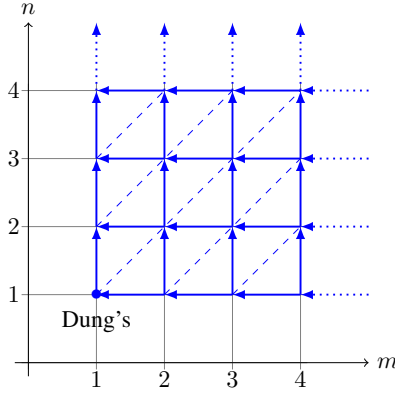\mathtt{d}_n^m(X) &\supseteq \mathtt{d}_{n+1}^m(X) & (16)
\end{aligned}
$$

Figure 5: Depiction of the partial order $\rhd$ over the set of all graded defense functions (with $0 < m, n \in \mathbb{N}$). The horizontal and vertical axes consist of the values of m and, respectively, n. Arrows go from 'weaker' to 'stronger' defense functions. Dashed lines (diagonals) denote incomparability. The point $m = n = 1$ denotes the position that Dung's characteristic function occupies in the ordering.

Proof. Recall the definition of the neutrality function (Definition 6). To establish (14) it suffices to notice that the property expresses the contrapositive of the following statement: if there exist at least $\ell + 1$ attackers in X then there exist at least ℓ attackers in X. Property (15) then follows directly by (14) above and (11) (Fact 4), through the following series of relations:

$$\mathtt{d}^m_n(X) = \mathtt{n}_m\big(\mathtt{n}_n(X)\big)$$
$$\subseteq \mathtt{n}_{m+1}\big(\mathtt{n}_n(X)\big) = \mathtt{d}^{m+1}_n(X).$$

A similar argument applies to establish (16), which follows by (14) above, (11), and the antitonicity of $\mathtt{n}$ (Fact 4):

$$\mathtt{d}^m_n(X) = \mathtt{n}_m\big(\mathtt{n}_n(X)\big)$$
$$\supseteq \mathtt{n}_m\big(\mathtt{n}_{n+1}(X)\big) = \mathtt{d}^m_{n+1}(X).$$

This completes the proof. $\qquad\square$

Intuitively, (14) states that the set of arguments attacked by at most ℓ arguments in X is included in the set of arguments attacked by at most $\ell + 1$ arguments in X. This establishes an ordering, in terms of logical strength, among the values of different neutrality functions: the lower is ℓ the stricter is the value of $\mathtt{n}_l$ applied to a same set of arguments X. Properties (15) and (16) then follow by combining this simple fact with the fact that $\mathtt{d}^m_n$ is the composition of $\mathtt{n}_m$ with $\mathtt{n}_n$ (11).

3.4. Comparing Graded Defense and Neutrality Functions

Fact 6 provides ground for a natural way in which different graded defense and neutrality functions can be ordered as their parameters m and n vary. The choice of
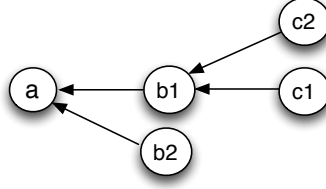
18

Figure 6: Framework of Example 8

these parameters determines the logical strength of different 'types' or 'standards' of conflict-freeness, which is based on neutrality, and acceptability, which is based on defense.

In light of Fact 6, comparing different neutrality functions is straightforward. Any relaxation on the requirement that no argument in a set be attacked by other arguments in that set leads to weaker forms of conflict-freeness. For any $X \subseteq \mathcal{A}$, $\mathtt{n}_\ell(X) \subseteq \mathtt{n}_k(X)$ for k and ℓ positive integers whenever $\ell \leq k$. So neutrality functions can simply be ordered linearly like natural numbers, with lower numbers denoting 'stronger' forms of neutrality and hence conflict-freeness.

The ordering of defense functions is more interesting, as these functions are parameterized by two integers:

Definition 7. $\mathtt{d}_{\underset{n}{m}} \rhd \mathtt{d}_{\underset{t}{s}}$ *(to be read "is at least as strong as") iff for any* $X \subseteq \mathcal{A}$, $\mathtt{d}_{\underset{n}{m}}(X) \subseteq \mathtt{d}_{\underset{t}{s}}(X)$, *with* m, n, s, t *positive integers.*

Relation $\rhd$ orders the set of all graded defense functions in a well-behaved manner:

Fact 7. *Let* $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ *be an AF, and let* $\rhd$ *be defined as above. Then:*

(i) $\mathtt{d}_{\underset{n}{m}} \rhd \mathtt{d}_{\underset{t}{s}}$ *iff* $m \leq s$ *and* $t \leq n$;

(ii) Relation $\rhd$ *is a* partial order, *i.e., reflexive, antisymmetric and transitive.*

Proof. $\boxed{(i)}$ is a direct consequence of Fact 6. $\boxed{(ii)}$ follows directly from how relation $\rhd$ is defined and the properties of set inclusion. $\square$

The relation is depicted in its generality in Figure 5. Expressions $\mathtt{d}_{\underset{n}{m}} \rhd \mathtt{d}_{\underset{t}{s}}$ may be read as follows: 'being mn-defended is *weakly preferable over* being st-defended' or 'the mn-defense function is *at least as strong as* the st-defense function'. Intuitively, the partial order $\rhd$ uses logical strength as a way to order graded defense functions. This equates with the intuition that if an argument meets a demanding standard of defense it also meets a less demanding one.

Example 8. *Referring to the framework in Figure 6, we illustrate Formula* (15) *and Fact 7:* $\mathtt{d}_{\underset{1}{1}}(\{c1, c2\}) = \{c1, c2, b2\} \subseteq \mathtt{d}_{\underset{1}{2}}(\{c1, c2\}) = \{c1, c2, a, b2\} \subseteq \mathtt{d}_{\underset{1}{3}}(\{c1, c2\}) = \{c1, c2, a, b1, b2\}$. *We also illustrate Formula* (16) *and Fact 7 with reference to Figure 4ii):* $\mathtt{d}_{\underset{1}{1}}(\{d2, c2, a2\}) = \{d2, c2, a2\} \supseteq \mathtt{d}_{\underset{2}{1}}(\{d2, c2, a2\}) = \{d2, c2, a2\} \supseteq \mathtt{d}_{\underset{3}{1}}(\{d2, c2, a2\})$ $= \{d2, c2\}$.

19

3.4.1. On the Partiality of $\rhd$

As the relation $\rhd$ is a partial order, some defense functions may be incomparable (see Figure 5) and this, we claim, is intuitive. By way of example, consider Dung's defense $\mathtt{d}_1$. This standard of defense is strengthened by $\mathtt{d}_{1\atop 1}$ (higher n parameter) and weakened by $\mathtt{d}_{2\atop 1}$ (higher m parameter). But under the definition of $\rhd$, $\mathtt{d}_{2\atop 2}$ defines a standard of defense which is incomparable with respect to $\mathtt{d}_{1\atop 1}$: it demands more defenders per attacker, but tolerates more attackers that are not counter-attacked to the desired level. In general, incomparability arises every time the parameters of the functions do not meet the condition $m \leq s$ and $t \leq n$ of Fact 7.

It should be clear, however, that the partial order $\rhd$ over graded defense functions could be further refined to a total order by resolving incomparability. This can be done in two ways: by either giving priority to parameter m or to parameter n. For example, if a set of arguments is mn-defended and another one is st-defended, where $m < s$ and $n < t$ (i.e., they are incomparable w.r.t. $\rhd$) then the first one can be stipulated to be more strongly defended because it is less tolerant with respect to the failure of defense. Therefore, for $m < s$ and $n < t$, belonging to $\mathtt{d}_{n}^{m}(X)$ is 'better' than belonging to $\mathtt{d}_{t}^{s}(X)$. One could then redefine $\rhd$ as follows: $\mathtt{d}_{n}^{m} \rhd \mathtt{d}_{t}^{s}$ iff either $m < s$, or $m = s$ and $n \geq t$. This yields a lexicographic order over graded defense functions giving priority to the m parameter over the n parameter. We do not investigate such refinements further in this paper.

4. Graded Semantics for Abstract Argumentation

By means of the graded defense and neutrality functions, Dung's notions of acceptability and conflict-freeness can be generalised to graded variants in a natural way. A set of arguments X is said to be conflict-free at grade ℓ (or, ℓ-conflict-free) whenever none of its arguments is attacked by at least ℓ arguments in X. A set of arguments X is said to be acceptable at grade mn (or, mn-acceptable) whenever all of its arguments are such that at most $m - 1$ of their attackers are not counter-attacked by at least n arguments in X. A graded notion of admissibility follows (ℓ-conflict-freeness plus mn-acceptability) and we thereby obtain graded variants of all the main admissibility-based semantics, which are simply Dung's standard semantics based on graded admissibility instead of standard admissibility. The first part of this section formally defines graded semantics. The rest of the section then develops a core theory of graded semantics. In the tradition of abstract argumentation, our results focus on the central questions of the existence and construction of graded extensions.

4.1. Graded Generalisation of Dung's Semantics

We are now in the position to generalise Definition 4 as follows:

Definition 8 (Graded Extensions)**.** *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF, $X \subseteq \mathcal{A}$, and ℓ, m and n be positive integers. Graded extensions are defined as in Table 2. We may write $\mathtt{adm}_{lmn}(\Delta)$, $\mathtt{prf}_{lmn}(\Delta)$ and $\mathtt{stb}_{lmn}(\Delta)$ to denote, respectively, the set of lmn-admissible, lmn-preferred. and lmn-stable extensions of Δ, and $\mathtt{grn}_{lmn}(\Delta)$ to denote the lmn-grounded extension of Δ. Finally, for an extension type $S \in$*

20

X is ℓ-conflict-free in Δ	iff	$X \subseteq \mathtt{n}_\ell(X)$
X is mn-self-defended	iff	$X \subseteq \mathtt{d}\underset{n}{m}(X)$
X is ℓmn-admissible in Δ	iff	$X \subseteq \mathtt{n}_\ell(X)$ and $X \subseteq \mathtt{d}\underset{n}{m}(X)$
X is an ℓmn-complete extension of Δ	iff	$X \subseteq \mathtt{n}_\ell(X)$ and $X = \mathtt{d}\underset{n}{m}(X)$
X is an ℓmn-stable extension of Δ	iff	$X = n_n(X) = n_m(X) \subseteq n_l(X)$
X is the ℓmn-grounded extension of Δ	iff	X is the smallest ℓmn-complete ext. of Δ
X is an ℓmn-preferred extension of Δ	iff	X is a largest ℓmn-complete ext. of Δ

Table 2: Graded generalizations of standard argumentation theory notions from [4].

$\{grounded, preferred, stable\}$, *we say that* $a \in \mathcal{A}$ *is* credulously *justified w.r.t.* ℓmn-S *if* $a \in \bigcup S_{\ell mn}(\Delta)$; *and* sceptically *justified w.r.t.* ℓmn-S *if* $a \in \bigcap S_{\ell mn}(\Delta)$ *Henceforth we assume the sceptical definition when referring to an argument simply as being*
530 justified.

The definition deserves some comment. Note first of all that when $l = m = n = 1$, we recover the standard definition of conflict-freeness, admissibility and extensions (Definition 4), which we henceforth refer to as 'Dung conflict-freeness' and 'Dung admissibility' and 'Dung extensions'. The key notion is graded admissibility, which is
535 obtained by parameterizing the conflict-freeness requirement by ℓ — i.e., $X \subseteq \mathtt{n}_\ell(X)$ —, and parameterizing the self-defense requirement by m and n — i.e., $X \subseteq \mathtt{d}\underset{n}{m}(X)$. The remaining graded semantics are defined by extending graded admissability in exactly the same way in which Dung admissibility is extended to define the standard Dung semantics. So, a graded complete extension, with parameters ℓ, m and n, is a fixpoint
540 of $\mathtt{d}\underset{n}{m}$, which is also ℓ-conflict-free, the graded grounded extension is the smallest ℓmn-complete extension, and the graded preferred extensions are the largest ℓmn-complete extensions. Finally, a graded stable extension, with parameters ℓ, m and n, is a fixpoint of $n_n(X)$ and $n_m(X)$ (and therefore of $\mathtt{d}\underset{n}{m}$), which is also ℓ-conflict-free. Constructive existence results for these semantics are provided in the next section.

545 Each graded extension type should then be interpreted as a class of weakenings and strengthenings of its standard Dung counterpart. For example: Dung complete extensions are strengthened by $11n$-complete extensions, with $n > 1$, which require a higher number of defenders for each attacked argument (that is, the requirements for acceptability are strengthened); and are weakened by $\ell 11$-complete extensions, with
550 $\ell > 1$, which tolerate a higher level of internal conflict (that is, weakening the conflict-freeness requirement), or by $1m1$-complete extensions, with $m > 1$, which tolerate a higher level of undefended arguments (that is, weakening the acceptability requirement). So for each Dung extension type, we now have an ordered family of extensions incorporating a form of graduality.

555 *4.2. Fixpoint Construction for Graded Exensions*

We proceed as in the standard case (cf. Section 2). The basic idea is as follows: given a graded admissible set, we show that, and under what assumptions on the parameters ℓ, m and n, this can be expanded into a graded complete set through a process of fixpoint approximation.

Fix a framework Δ and take a set X such that $X \subseteq \mathrm{d}_n^m(X)$ and $X \subseteq \mathrm{n}_\ell(X)$ (that is, an ℓmn-admissible set). By iterating d_n^m, consider the stream of sets

$$\mathrm{d}_n^{m0}(X), \mathrm{d}_n^{m1}(X), \ldots \tag{17}$$

and the stream of sets

$$\mathrm{n}_n(\mathrm{d}_n^{m0}(X)), \mathrm{n}_n(\mathrm{d}_n^{m1}(X)), \ldots \tag{18}$$

which, by Fact 4, is equivalent to the stream

$$\mathrm{d}_m^{n0}(\mathrm{n}_n(X)), \mathrm{d}_m^{n1}(\mathrm{n}_n(X)), \ldots. \tag{19}$$

By the above assumptions on X, and since d_n^m is monotonic and n_n is antitonic (cf. Fact 4), the stream in (17) is non-decreasing and the stream in (19) is non-increasing, with respect to set inclusion. In finite attack graphs, these streams must therefore stabilize reaching a limit at iteration $|\mathcal{A}| + 1$. In infinite but finitary attack graphs, we will see that the limit can be reached at ω. We will also see (Lemma 4) that the limits of these streams correspond to the *smallest* fixpoint of d_n^m *containing* X and, respectively, the *largest* fixpoint of d_m^n which is *contained* in $\mathrm{n}_n(X)$.[17] We denote the first one by $\mathrm{lfp}_X.\mathrm{d}_n^m$ and the second one by $\mathrm{gfp}_X.\mathrm{d}_m^n$. Intuitively, the two sets denote the smallest superset of X which is equal to the set of arguments it mn-defends and, respectively, the largest set whose arguments are not attacked by at least n arguments in X and which is equal to the set of arguments it nm-defends. The above construction is illustrated in Figure 7 below. We prove now its correctness showing that the limits of the streams in (17) and (19) correspond indeed to the desired fixpoints.

First of all the following important lemma shows under what conditions on m and n graded conflict-freeness can be preserved by the above process of iteration of the defense function.[18]

Lemma 3. *Let Δ be a finitary attack graph, $X \subseteq \mathcal{A}$ be such that $X \subseteq \mathrm{d}_n^m(X)$ and $X \subseteq \mathrm{n}_m(X)$, and m, n be two positive integers such that $n \geq m$. Then for any k s.t. $0 \leq k < \omega$,*
$$X \subseteq \mathrm{d}_n^{mk}(X) \subseteq \mathrm{n}_m(\mathrm{d}_n^{mk}(X)) \subseteq \mathrm{n}_m(X).$$

Proof. $\boxed{X \subseteq \mathrm{d}_n^{mk}(X)}$ holds by the assumption that X is admissible and by the monotonicity of d_n^m (Fact 4). $\boxed{\mathrm{d}_n^{mk}(X) \subseteq \mathrm{n}_m(\mathrm{d}_n^{mk}(X))}$ is proven by induction over k. The $\boxed{\text{base case}}$ holds by assumption as $\mathrm{d}_n^{m0}(X) = X$ is mmn-admissible and therefore

---

[17] Notice the reversal in the parameters m and n due to the fact that $\mathrm{n}_n(\mathrm{d}_n^m(X)) = \mathrm{d}_m^n(\mathrm{n}_n(X))$, a direct consequence of Fact 4.

[18] In the following two lemmas we handle only parameters m and n directly. This, we will see, is sufficient to establish results concerning also parameter ℓ later (Theorem 2).
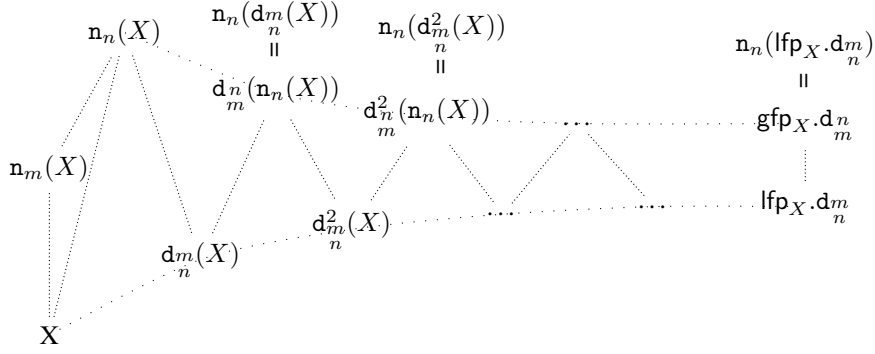
Figure 7: The two streams of (17) and (19) under the assumption that X is mmn-admissible and the two integers n and m are such that $n \geq m$. Position with respect to the horizontal axis indicates the number of iterations, and positions with respect to the vertical axis indicates set theoretic inclusion. Cf. Figure 2 depicts the same behaviour for the standard, non-graded, case.

m-conflict-free. For the $\boxed{\text{induction step}}$ assume (IH) that $\mathsf{d}_m^k(X) \subseteq \mathsf{n}_m(\mathsf{d}_m^k(X))$ (that

is, the kth step is m-conflict-free). We show that $\mathsf{d}_m^{k+1}(X) \subseteq \mathsf{n}_m(\mathsf{d}_m^{k+1}(X))$. Suppose towards a contradiction that is not the case. Then there exists x and $y_1, \ldots, y_m$ in set $\mathsf{d}_m^{k+1}(X) = \mathsf{d}_m(\mathsf{d}_m^k(X))$ such that $x \leftarrow y_i$ for $1 \leq i \leq m$. Since $n \geq m$ by assumption, by the definition of d_m there exist $z_1, \ldots, z_n \in \mathsf{d}_m(X)$ such that $y_\ell \leftarrow z_i$ for $1 \leq i \leq n$. That is, at least one among all y_i's (w.l.o.g. assumed to be y_m) is attacked by at least n arguments in $\mathsf{d}_m(X)$. But from this and the assumption that $n \geq m$, there exist $w_1, \ldots, w_n \in \mathsf{d}_m(X)$ such that $z_n \leftarrow w_i$ for $1 \leq i \leq n$. That is, at least one among all z_i's (w.l.o.g. assumed to be z_n) is also attacked by at least n arguments in $\mathsf{d}_m(X)$. Now since $n \geq m$ we conclude that $\mathsf{d}_m(X)$ is not m-conflict-free, against IH. $\boxed{\mathsf{n}_m(\mathsf{d}_m^k(X)) \subseteq \mathsf{n}_m(X)}$ follows from the first claim by the antitonicity of n_m (Fact 4). $\qquad\square$

That is, each stage $\mathsf{d}_m^k(X)$ in the stream of iteration of the graded defense function from set X is an m-conflict-free set.

Lemma 4. *Let Δ be a finitary attack graph and $X \subseteq \mathcal{A}$ be such that $X \subseteq \mathsf{d}_m(X)$ and $X \subseteq \mathsf{n}_m(X)$, and m, n be two positive integers such that $n \geq m$:*

$$\mathsf{lfp}_X.\mathsf{d}_m = \bigcup_{0 \leq k < \omega} \mathsf{d}_m^k(X) \tag{20}$$

$$\mathsf{gfp}_X.\mathsf{d}_m = \bigcap_{0 \leq k < \omega} \mathsf{d}_n^k(\mathsf{n}_n(X)) \tag{21}$$

Proof. $\boxed{(20)}$ $\boxed{\text{First}}$, we prove that $\bigcup_{0 \leq n < \omega} \mathsf{d}_m^k(X)$ is a fixpoint by the following se-

23

ries of equations:

$$\mathtt{d}_{\substack{m\\n}}\left(\bigcup_{0\le k<\omega}\mathtt{d}_{\substack{m\\n}}^{k}(X)\right)=\bigcup_{0\le k<\omega}\mathtt{d}_{\substack{m\\n}}(\mathtt{d}_{\substack{m\\n}}^{k}(X))$$

$$=\bigcup_{0\le k<\omega}\mathtt{d}_{\substack{m\\n}}^{k}(X)$$

where the first equation holds by the ω-continuity of $\mathtt{d}_{\substack{m\\n}}$ (Fact 5) and the second by the fact that, since X is ℓmn-admissible by assumption and $\mathtt{d}_{\substack{m\\n}}$ is monotonic (Fact 4) we have that: $X=\mathtt{d}_{\substack{m\\n}}^{0}(X)\subseteq\mathtt{d}_{\substack{m\\n}}^{k}(X)\subseteq\mathtt{d}_{\substack{m\\n}}^{k+1}(X)$ for any k s.t. $0\le k<\omega$. $\boxed{\text{Second}}$, we prove that $\bigcup_{0\le k<\omega}\mathtt{d}_{\substack{m\\n}}^{k}(X)$ is indeed the least fixpoint containing X. Suppose, towards a contradiction that there exists Y s.t.: $X\subset Y=\mathtt{d}_{\substack{m\\n}}^{k}(Y)\subset\bigcup_{0\le k<\omega}\mathtt{d}_{\substack{m\\n}}^{k}(X)$. It follows that $X\subset Y=\mathtt{d}_{\substack{m\\n}}^{k}(Y)\subset\mathtt{d}_{\substack{m\\n}}^{k}(X)$ for some k s.t. $0\le k<\omega$. But, by the monotonicity of $\mathtt{d}_{\substack{m\\n}}$ (Fact 4), we have that $\mathtt{d}_{\substack{m\\n}}^{k}(X)\subseteq\mathtt{d}_{\substack{m\\n}}^{k}(Y)$. Contradiction.

$\boxed{(21)}$ The proof is similar to the previous case. $\boxed{\text{First}}$, we prove that $\bigcap_{0\le k<\omega}\mathtt{d}_{\substack{n\\m}}^{k}(\mathtt{n}_{n}(X))$ is a fixpoint, through the series of equations

$$\mathtt{d}_{\substack{n\\m}}\left(\bigcap_{0\le k<\omega}\mathtt{d}_{\substack{n\\m}}^{k}(\mathtt{n}_{n}(X))\right)=\bigcap_{0\le k<\omega}\mathtt{d}_{\substack{n\\m}}(\mathtt{d}_{\substack{n\\m}}^{k}(\mathtt{n}_{n}(X)))$$

$$=\bigcap_{0\le n<\omega}\mathtt{d}_{\substack{n\\m}}^{k}(\mathtt{n}_{n}(X))$$

which hold by the ω-continuity of $\mathtt{d}_{\substack{n\\m}}$ (Fact 5) and by the fact that, since X is mmn-admissible by assumption, and $\mathtt{d}_{\substack{n\\m}}$ is monotonic (Fact 4), $\mathtt{n}_{n}(X)=\mathtt{d}_{\substack{n\\m}}^{0}(\mathtt{n}_{n}(X))\supseteq\mathtt{d}_{\substack{n\\m}}^{k}(\mathtt{n}_{n}(X))$ for any k s.t. $0\le k<\omega$. The latter property holds because X is assumed to be such that $X\subseteq\mathtt{d}_{\substack{m}}(X)$. By the antitonicity of $\mathtt{n}_{n}$ and the interdefinability of $\mathtt{d}_{\substack{m\\n}}$ as $\mathtt{n}_{n}\circ\mathtt{n}_{m}$ (Fact 4) we therefore have that $\mathtt{n}_{n}(\mathtt{d}_{\substack{m\\n}}(X))=\mathtt{d}_{\substack{n\\m}}(\mathtt{n}_{n}(X))\subseteq\mathtt{n}_{n}(X)$ from which it follows that $\mathtt{d}_{\substack{n\\m}}(\mathtt{n}_{n}(X))\subseteq\mathtt{n}_{n}(X)$. A simple induction on k then establishes the claim. $\boxed{\text{Second}}$, it remains to be proven that $\bigcap_{0\le k<\omega}\mathtt{d}_{\substack{n\\m}}^{k}(\mathtt{n}_{n}(X))$ is indeed the largest fixpoint of $\mathtt{d}_{\substack{n\\m}}$ contained in $\mathtt{n}_{n}(X)$. Like in the previous case we proceed towards a contradiction. Suppose there exists Y s.t.: $\bigcap_{0\le k<\omega}\mathtt{d}_{\substack{n\\m}}^{k}(\mathtt{n}_{n}(X))=\bigcap_{0\le k<\omega}\mathtt{n}_{n}(\mathtt{d}_{\substack{m\\n}}^{k}(X))\subset Y=\mathtt{d}_{\substack{n\\m}}(Y)\subseteq\mathtt{n}_{n}(X)$. There must therefore exist an integer k such that, as a consequence of Lemma 3, $X\subseteq\mathtt{d}_{\substack{m\\n}}^{k}(X)\subseteq\mathtt{n}_{n}(\mathtt{d}_{\substack{m\\n}}^{k}(X))\subset Y$. By the antitonicity of $\mathtt{n}_{n}$ and again by the interdefinability of $\mathtt{d}$ and $\mathtt{n}$ (Fact 4), and since Y is taken to be a fixpoint of $\mathtt{d}_{\substack{n\\m}}$, it follows that $\mathtt{n}_{n}(Y)=\mathtt{n}_{n}(\mathtt{d}_{\substack{n\\m}}(Y))=\mathtt{d}_{\substack{m\\n}}(\mathtt{n}_{n}(Y))$. Then, from the fact that $\mathtt{n}_{n}(\mathtt{d}_{\substack{m\\n}}^{k}(X))\subset Y$, and the fact that $\mathtt{n}_{m}$ is antitonic (Fact 4) we have that $\mathtt{n}_{n}(Y)=\mathtt{n}_{m}(\mathtt{n}_{n}(\mathtt{d}_{\substack{m\\n}}^{k}(X)))=\mathtt{d}_{\substack{m\\n}}(\mathtt{d}_{\substack{n\\m}}^{k}(X))$ which is a subset of $\bigcup_{0\le k<\omega}\mathtt{d}_{\substack{m\\n}}^{k}(X)$. So we have that $\mathtt{n}_{n}(Y)$ is also a fixpoint of $\mathtt{d}_{\substack{m\\n}}$, it contains X and it is included in

24

$\bigcup_{0 \leq k < \omega} \mathsf{d}_{\underset{n}{m}}^k(X)$, which, by the previous claim, we know to be the smallest fixpoint of d_Δ containing X. Contradiction. □

The content of Lemmas 3 and 4 underpinning the construction of the two fixpoints is depicted in Figure 7. Notice that like in the standard case of d_1 and n_1 the two streams can be generated by the indefinite iteration of the application of n_n followed by n_m. The fact that $n \geq m$ is assumed, guarantees that each set of arguments $\mathsf{d}_{\underset{n}{m}}^k(X)$ in the lower stream remains included in the set of arguments towards which it is m-neutral; that is, it remains m-conflict fee (and a fortiori n-conflict-free).

Remark 2. *Like for the proof of Lemma 2, finitariness plays an essential role in the proof of Lemma 4. However, also in this case the assumption can be lifted and the proof could proceed using transfinite induction (cf. Remark 1).*

4.2.1. Constructing Graded Complete and Grounded Extensions

We now move on to show how the above results provide constructive proofs of existence of graded semantics. We will focus on finitary graphs, but it should be clear that the non-finitary case can be handled by ordinal induction.[19]

Theorem 2. *Let Δ be a finitary AF, $X \subseteq \mathcal{A}$ be such that $X \subseteq \mathsf{n}_\ell(X)$ and $X \subseteq \mathsf{d}m$, with ℓ, m and n positive integers such that $n \geq m$ and $\ell \geq m$. Then $\bigcup_{0 \leq k < \omega} \mathsf{d}_{\underset{n}{m}}^k(X)$ is the smallest ℓmn-complete extension containing X.*

Proof. By Lemma 4, we know that $\bigcup_{0 \leq k < \omega} \mathsf{d}_{\underset{n}{m}}^k(X) = \mathsf{lfp}_X.\mathsf{d}_{\underset{n}{m}}$, that is, the smallest fixpoint of $\mathsf{d}_{\underset{n}{m}}$ that contains X. By Lemma 3 we know that this set is m-conflict-free, that is, $\bigcup_{0 \leq k < \omega} \mathsf{d}_{\underset{n}{m}}^k(X) \subseteq \mathsf{n}_m(\bigcup_{0 \leq k < \omega} \mathsf{d}_{\underset{n}{m}}^k(X))$. Since $\ell \geq m$ by assumption, by Fact 6 and the transitivity of $\subseteq$ we obtain that $\bigcup_{0 \leq k < \omega} \mathsf{d}_{\underset{n}{m}}^k(X) \subseteq \mathsf{n}_\ell(\bigcup_{0 \leq k < \omega} \mathsf{d}_{\underset{n}{m}}^k(X))$. It therefore follows that $\bigcup_{0 \leq k < \omega} \mathsf{d}_{\underset{n}{m}}^k(X)$ is a fixpoint of $\mathsf{d}_{\underset{n}{m}}$, it is ℓ-conflict-free and it is the smallest such set containing X. As claimed, it is therefore the smallest ℓmn-complete extension (Definition 8) containing X. □

Corollary 1. *Let Δ be a finitary AF, $n \geq m$ and $\ell \geq m$. Then $\bigcup_{0 \leq k < \omega} \mathsf{d}_{\underset{n}{m}}^k(\emptyset)$ is the ℓmn-grounded extension of Δ.*

So given an ℓmn admissible set ($n \geq m$ and $\ell \geq m$), the smallest ℓmn-complete extension containing X is the least fixpoint of $\mathsf{d}_{\underset{n}{m}}$ which contains X ($\mathsf{lfp}_X.\mathsf{d}_{\underset{n}{m}}$), and the ℓmn-grounded extension is simply the smallest fixpoint of $\mathsf{d}_{\underset{n}{m}}$ ($\mathsf{lfp}.\mathsf{d}_{\underset{n}{m}}$).

4.2.2. Constructing Graded Preferred Extensions

Theorem 2 implies that if we choose a 'large enough' ℓmn-admissible set X, in the sense that such a set can reach through a chain of attacks any other argument, then the indefinite iteration of $\mathsf{d}_{\underset{n}{m}}$ will yield an ℓmn-preferred extension (when $n \geq m$ and $\ell \geq m$).

---

[19]Cf. Remark 1.

Fact 8. *Let Δ be a finitary AF, $X \subseteq \mathcal{A}$ be such that $X \subseteq \mathtt{n}_\ell(X)$ and $X \subseteq \mathtt{d}_{\substack{m \\ n}}(X)$, and $n \geq m$ and $\ell \geq m$. Assume furthermore that $\mathcal{A} \subseteq \{a \mid \exists b \in X : a \leftarrow^+ b\}$. Then $\bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X)$ is the ℓmn-preferred extension of Δ that contains X.*

Proof. By Theorem 2 we know that $\bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X)$ is the smallest ℓmn-complete extension containing X. It needs to be shown that there exists no ℓmn-admissible set Y such that $X \subset Y$. Suppose, towards a contradiction, that this is the case. Then $X \subseteq \bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X) \subset Y \subseteq \mathtt{d}_{\substack{m \\ n}}(Y) \subseteq \{a \mid \exists b \in X : a \leftarrow^+ b\}$. It follows that there exists $x \in Y \subseteq \mathtt{d}_{\substack{m \\ n}}(Y)$ such that $x \notin \mathtt{d}_{\substack{k \\ m \\ n}}(X)$ for some positive integer k. By assumption, there exists a finite path of attacks from X to x. So let X_1 be the smallest set of arguments that mn-defends x, X_2 the smallest set of arguments that mn-defends the arguments in X_1 and so on. It follows there exists a j such that $X_j \subseteq X$, otherwise x would not be mn-defended. It therefore follows that $x \in \bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X)$. Contradiction. $\square$

4.2.3. Constructing Graded Stable Extensions

We finally arrive at a characterization of graded stable extensions as limits of streams generated by the graded neutrality function:

Fact 9. *Let Δ be a finitary AF, $X \subseteq \mathcal{A}$ be such that $X \subseteq \mathtt{n}_\ell(X)$ and $X \subseteq \mathtt{d}_{\substack{m \\ n}}(X)$, and $n \geq m$ and $\ell \geq m$. Then, $\bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X)$ is the smallest ℓmn-stable extension containing X if and only if $\bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X) = \bigcap_{0 \leq k < \omega} \mathtt{n}_n(\mathtt{d}_{\substack{k \\ m \\ n}}(X))$.*

Proof. $\boxed{\Leftarrow}$ Assume $\bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X) = \bigcap_{0 \leq k < \omega} \mathtt{n}_n(\mathtt{d}_{\substack{k \\ m \\ n}}(X))$. It therefore follows that $\bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X) = \mathtt{n}_n(\bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X))$. We conclude that $\bigcup_{0 \leq k < \omega} \mathtt{d}_{\substack{k \\ m \\ n}}(X)$ is a fixpoint of $\mathtt{n}_n$, that is by Definition 8, an ℓmn-stable extension. $\boxed{\Rightarrow}$ Straightforward. $\square$

So, as in the standard case, graded stable extensions are the results of the convergence of the upper and lower streams of iteration of the graded defense function (cf. Figure 7). As in the standard case, such convergence is not guaranteed in general.

4.3. On the constraints $n \geq m$ and $\ell \geq m$

Theorem 2 assumed that the three parameters of graded admissibility ℓ, m and n are in the relation $n \geq m$ and $\ell \geq m$. Notice that Dung's standard semantics trivially meets this constraint with $\ell = m = n = 1$. This assumption plays a crucial role in the proofs of the above results and one can in fact show that graded semantics for a choice of parameters failing the constraint may not exist in some frameworks.[20]

---

[20]This is a situation fully analogous to the case of stable extensions in the standard Dung framework, where they are not always guaranteed to exist.

Example 9. *Let Δ be the AF consisting of the arguments and attacks in Figure 4iii). Then Δ has no 221-grounded extension. To establish this let us try to construct such an extension with the construction of Theorem 2:*

$$\mathrm{d}_{\underset{1}{2}}(\emptyset) = \{b_3, c_3, d_3, e_3\}, \mathrm{d}_{\underset{1}{2}}(\{b_3, c_3, d_3, e_3\}) = \{a_3, b_3, c_3, d_3, e_3\}$$

So the whole set of arguments $\mathcal{A}$ in that framework constitutes the smallest fixpoint of $\mathrm{d}_{\underset{1}{2}}$. Such a set is, however, not 2-conflict-free, that is $\mathcal{A} \not\subseteq \mathrm{n}_2(\mathcal{A})$. In fact no 221-complete extensions exist in this framework, as these would have to include the set of unattacked arguments $\{d_3, e_3\}$ which, it is easy to see, 21-defends all arguments.[21]

Intuitively, the constraint imposes two properties on graded admissibility: first, that the level m of failure of defense which we are willing to tolerate, should not exceed the number n of counter-attackers we want for such a defense to hold; second, that such a level m of failure of defense, should not exceed the level ℓ of conflict-freeness that we are willing to tolerate. So to guarantee existence of a graded semantics tolerating arguments for which m counter-attacks fail, one has to set the number n of required counter-attackers higher or equal to m; and similarly, one has to set the level of tolerance ℓ to internal attacks at least as high as m.[22]

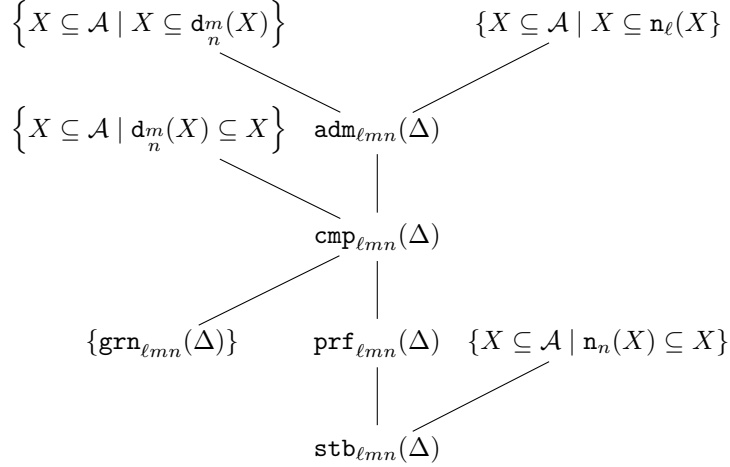## 4.4. Basic Properties of Graded Semantics

A direct consequence of Definition 8 is that graded extensions are in the same logical relations as standard Dung's extensions:

Fact 10. *For any AF $\Delta = \langle \mathcal{A}, \rightarrow \rangle$, and integers ℓ, m and n, graded semantics are*

---

[21]See Example 10 later for another such example.

[22]Consider a 232-admissible extension X (where in violation of the constraints, $l = 2, m = 3, n = 2$), and some $x \in X$ such that x is attacked by two undefended attackers, in keeping with the m parameter indicating that x can be considered acceptable up to a cumulative threshold of two fully acceptable (i.e., two undefended) attackers of x. This level of tolerance with respect to the acceptability of x is then at odds with the more restrictive toleration of the co-acceptability of x with a maximum of one attacker $y \in X$ on x (as indicated by $l = 2$). Moreover, the fact that a third attacker z of x need only be counter-attacked by two arguments (as indicated by $n = 2$), means that one can still (given $m = 3$) consider z to be an acceptable argument by the above reasoning applied to x, which in turn implies that the tolerated cumulative threshold on attackers of x has been exceeded. Precisely when this sort of mismatch between levels of tolerance of acceptability and co-acceptability occurs, graded semantics may fail to exist.

related according to the following diagram:

$$\left\{X \subseteq \mathcal{A} \mid X \subseteq \mathtt{d}_n^m(X)\right\} \qquad\qquad \{X \subseteq \mathcal{A} \mid X \subseteq \mathtt{n}_\ell(X)\}$$

$$\left\{X \subseteq \mathcal{A} \mid \mathtt{d}_n^m(X) \subseteq X\right\} \quad \mathtt{adm}_{\ell mn}(\Delta)$$

$$\mathtt{cmp}_{\ell mn}(\Delta)$$

$$\{\mathtt{grn}_{\ell mn}(\Delta)\} \qquad \mathtt{prf}_{\ell mn}(\Delta) \quad \{X \subseteq \mathcal{A} \mid \mathtt{n}_n(X) \subseteq X\}$$

$$\mathtt{stb}_{\ell mn}(\Delta)$$

where two nodes are connected when the lower node in the pair is a subset ($\subseteq$) of the upper node in the pair.

So all ℓmn extensions are mn-admissible and ℓ-conflict-free; ℓmn-grounded, -stable and -preferred are all ℓmn-complete extensions; and ℓmn-stable extensions are ℓmn-preferred. We will further illustrate Definition 8, and the associated constructions, with a series of examples later in Section 4.5.

Fact 11. *For any AF Δ, and integers ℓ, m and n such that $n \geq m$ and $\ell \geq m$:*

$$\{x \mid \overline{x} = \emptyset\} \subseteq \mathtt{grn}_{\ell mn}(\Delta) \tag{22}$$

$$\bigcup S_{\ell mn}(\Delta) \subseteq \bigcup S_{\ell' m' n'}(\Delta) \tag{23}$$

$$\bigcap S_{\ell mn}(\Delta) \subseteq \bigcap S_{\ell' m' n'}(\Delta) \tag{24}$$

for $S \in \{grounded, preferred, stable\}$ and any ℓ', m', n' such that $n' \geq m'$ and $\ell' \geq m'$ and $\ell' \geq \ell$, $m' \geq m$ and $n' \leq n$.

Proof. $\boxed{(22)}$ It is easy to see that for any ℓ, m and n if $\overline{x} = \emptyset$ then $x \in \mathtt{d}_n^m(\emptyset)$ and therefore, by Definition 8 $x \in \mathtt{grn}_{\ell mn}(\Delta)$. $\boxed{(23) \& (24)}$ Both claims follow from Definition 8 by Fact 6. The constraint on ℓ', m' and n' is necessary to guarantee existence under such parameters. $\square$

4.5. Some Examples

Example 10. *Referring to Figure 8's AF below:*

$$\mathtt{d}_{\frac{1}{2}}(\emptyset) = \{d, f, g\} \, ; \mathtt{d}_{\frac{1}{2}}(\{d, f, g\}) = \{d, f, g\} \, .$$

28

So $\{d, f, g\}$ is the smallest fixpoint of $\mathrm{d}_{\substack{1 \\ 2}}$ and it is 1-conflict-free (as well as ℓ-conflict-free for every $\ell \geq 1$). It is therefore also the 112-grounded extension (as well as ℓ12-grounded extension for every $\ell \geq 1$) of the given framework. Consider now:

$$\begin{array}{ccc} \mathrm{d}_{\substack{1 \\ 2}}(\emptyset) & \mathrm{d}_{\substack{1 \\ 2}}(\{c, d, f, g\}) & \mathrm{d}_{\substack{1 \\ 2}}(\{a, b, c, d, f, g\}) \\ \| & \| & \| \\ \{c, d, f, g\} & \{a, b, c, d, f, g\} & \{a, b, c, d, f, g\} \end{array}$$

So $\{a, b, c, d, f, g\}$ is the smallest fixpoint of $\mathrm{d}_{\substack{1 \\ 2}}$. However, it is not 2-conflict-free (argument a in the set is attacked by two other arguments in the set) and therefore it is not a 221-grounded extension. To make it a graded grounded extension one has to tolerate more internal attacks, setting for instance $\ell = 3$. The set is indeed a 321-grounded extension.

Example 11. *Consider now the central graph (3-cycle) in Figure 1. We know that such a graph has no grounded extension and the only complete extension is $\emptyset$. But one can slightly relax the defence and conflict-free requirements to obtain a non-empty graded complete extension. Set $\ell = 2$, $m = 2$ and $n = 1$. We can construct the 221-grounded extension of this framework as follows:*

$$\mathrm{d}_{\substack{1 \\ 2}}(\emptyset) = \{a, b, c\} \, , \mathrm{d}_{\substack{1 \\ 2}}(\{a, b, c\}) = \{a, b, c\}$$

So, $\{a, b, c\}$ is the smallest fixpoint of the graded defense function $\mathrm{d}_{\substack{1 \\ 2}}$ and such a set is clearly 2-conflict-free (that is, $\{a, b, c\} \subseteq \mathrm{n}_2(\{a, b, c\})$). It is therefore also a 221-preferred and -stable extension.

5. Ranking Arguments by Graded Semantics

The theory developed in the previous sections offers a novel perspective on how arguments can be compared from an (abstract) argumentation theoretic point of view. In this section we describe two natural ways in which the theory of graded acceptability can be applied to define orderings on arguments: a 'contextual' way, whereby, given a fixed set of arguments, one can rank arguments by how well they are iteratively defended by the given set; and an 'absolute' way, whereby arguments are compared based on their acceptability under given graded semantics. The definitions introduced are illustrated by means of several examples in this section and later in Section 6.

5.1. Contextual Approach: Ranking by Quality of Defense

The same recursive principles underpinning the standard Dung semantics can be used to characterise how strongly the set of arguments defended by a given set, defends another set, and how the latter set defends yet another set, and so forth. That is to say, given a set X—the context—one can iteratively apply $\mathrm{d}_{\substack{m \\ n}}$ to X. Iterated graded defense can thus rank arguments with respect to how well a given set X defends them:

Definition 9. *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be a finitary AF and $X \subseteq \mathcal{A}$. For $a, b \in \mathcal{A}$, we define that a is 'at least as justified as' b w.r.t X as follows:*

$$a \succeq^X b \Longleftrightarrow \forall m, n > 0 \text{ IF } b \in \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{n}{m}}(X) \text{ THEN } a \in \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{n}{m}}(X)$$

The strict part $\succ_X$ of the above relation is defined in the obvious way.

As usual, $a \approx^X b$ denotes that $a \succeq^X b$ and $b \succeq^X a$. When we want to make the underlying AF explicit we use the heavier notation $a \succeq^X_\Delta b$.

The key intuition behind this definition is the following. Take two arguments a and b, and some fixed set X. Is it the case that every time b is defended through the iteration of some graded defense function, a also is? If that is the case, it means that (w.r.t. X) every standard of defense met by b is also met by a, but a may satisfy yet stronger ones (recall Fact 7). It is easy to see that $\succeq^X$ is a partial order, for any $X \subseteq \mathcal{A}$.

Example 12. *Let us rank, by iterated defense w.r.t $\emptyset$, the arguments in Figure 4i-iv). Applying Definition 9 we obtain the partial order shown in the Hasse diagram in Figure 4vii). Note that we assume one single AF Δ consisting only of the arguments and attacks shown in Figures 4i-iv). Under the standard Dung semantics, all arguments in $\bigcup_{i=1}^{4} Xi$ are in the iterated application of d_Δ to $\emptyset$ (i.e., in the grounded extension of Δ). However, we can now differentiate amongst these arguments. As expected the best arguments are those with no attackers. Second-best is $a2$ whose attackers are all counterattacked by two un-attacked arguments. Third-best is $a1$, since $a2 \in \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{2}{1}}(\emptyset)$ but $a1 \notin \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{2}{1}}(\emptyset)$ (and so $a1 \not\succeq^\emptyset a2$), but $\forall m, n$: if $a1 \in \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{n}{m}}(\emptyset)$ then $a2 \in \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{n}{m}}(\emptyset)$ (and so $a2 \succeq^\emptyset a1$). We then have that $a3$ and $a4$ are incomparable (recall Section 3.4.1). Formally, $a3 \in \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{3}{3}}(\emptyset)$ and $a3 \notin \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{2}{2}}(\emptyset)$, but $a4 \notin \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{3}{3}}(\emptyset)$ and $a4 \in \bigcup_{0 \leq k < \omega} \mathsf{d}^k_{\underset{2}{2}}(\emptyset)$. Critically, we can also differentiate amongst the rejected arguments (those not in the Dung grounded extension). Thus $b1, b3, c3, c4$, (each of which are attacked by one argument), are ranked above $b2, b4, d4$ (each of which are attacked by two arguments).*

5.2. Absolute Approach: Ranking by Quality of Justification

More generally, for a given semantics, we can rank the justification status of an argument with respect to a given framework, exactly as we did in Definition 9 for iterated graded defense:

Definition 10 (Ranking arguments by graded semantics)**.** *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF. For $a, b \in \mathcal{A}$, and for $S \in \{\ell mn\text{-grounded}, \ell mn\text{-stable}, \ell mn\text{-preferred}\}$:[23]*

$$a \succeq^S b \Longleftrightarrow \forall \ell, m, n > 0, \text{ IF } b \text{ is justified w.r.t. } S \text{ THEN } a \text{ is justified w.r.t. } S.$$

The strict part $\succ_S$ of the above relation is defined in the obvious way.

---

[23]Recall we are working with the sceptical notion of justifiability throughout the paper.
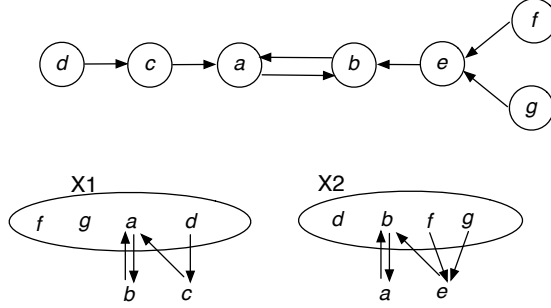
Figure 8: An AF with two preferred extensions

As usual, $a \approx^S b$ denotes that $a \succeq^S b$ and $b \succeq^S a$. Again, it is easy to see that $\succeq^S$ is a partial order. When we want to make the underlying AF explicit we use the heavier notation $a \succeq^S_\Delta b$.

765   We illustrate how the ranking of arguments by graded semantics can be applied to arbitrate between arguments that are credulously but not sceptically justified under Dung's standard semantics.

     Consider the AF Δ in Figure 8 that has two preferred extensions—$X1$ and $X2$— under the standard Dung semantics. This equates with $X1$ and $X2$ both being 111-

770 preferred extensions in the graded terminology. Hence a and b are credulously justified, while only f, g and d are sceptically justified. Typically, one would then rely on exogenously given preferences [6, 8] to arbitrate between such arguments. However, we can make use of the endogenously derived ranking of arguments yielded by our graded semantics to arbitrate amongst a and b. Intuitively, b is more strongly defended in $X2$

775 than a is defended in $X1$, and the graded semantics can make use of this information in the standard framework to arbitrate in favour of b over a. Specifically, only $X2$ is a 222-admissible (and hence a subset of a 222-preferred[24]) extension since only one attacker (a) of b is defended by strictly less than two arguments, whereas $X1$ is not a 222-admissible ($X1$ is not a subset of a 222-preferred) extension since both attackers of

780 a (b, c) are defended by strictly less than two arguments. Indeed, $\forall l, m, n \geq 0$: IF a is justified w.r.t. lmn-preferred THEN b is justified w.r.t. lmn-preferred, whereas *it is not the case* (as illustrated above) that $\forall l, m, n \geq 0$: IF b is justified w.r.t. lmn-preferred THEN a is justified w.r.t. lmn-preferred. Hence $b \succ^{lmn\text{-preferred}}_\Delta a$, so arbitrating in favour of b over a.

785   It is worth noting that the contextual and absolute approaches to argument ranking are related as follows.

Fact 12. *Let Δ be a finitary AF, $n \geq m$ and $\ell \geq m$. Then:*

$$a \succeq^\emptyset b \iff a \succeq^{lmn\text{-grounded}} b$$

---

[24]The 222-preferred extension is $X2 \cup \{a\}$.

31

Proof. The claim is a direct consequence of Corollary 1 and Definitions 9 and 10. $\quad\square$

Further properties of rankings based on graded semantics will be discussed in detail later (Section 7), in the context of other existing approaches to argument rankings in argumentation.

6. Instantiating Graded Semantics

We have thus far focussed on graded defense, acceptability and semantics for abstract argumentation frameworks. We now illustrate these notions by reference to instantiated frameworks. In particular, *ASPIC+* [8, 33] provides a general framework for specifying logical instantiations of Dung frameworks. It has been shown to formalise human orientated accounts of argumentation based reasoning that make use of Schemes and Critical Questions [25], and provide a dialectical characterisation of non-monotonic inference in Brewka's Preferred Subtheories [24] and Prioritised Default Logic [34] (in [8] and, respectively, [35]). In what follows we show how our graded semantics can be applied to both these instantiations.

6.1. Graded Semantics for Instantiations based on Schemes and Critical Questions

We use a well-established instantiated argumentation setting to illustrate the usefulness of the theory of graded acceptability for evaluating argument strength.

6.1.1. Schemes and Critical Questions

Schemes and Critical Questions (*SchCQ*) have been developed by the informal logic community, most notably by Walton [25], and capture stereotypical patterns of argument as deployed in epistemic and practical reasoning. For example, the *argument from expert opinion* scheme states that if E is an expert in domain D, and E states that S is true (false), and S is within domain D, then (presumably) S is true (false).

Echoing our motivation for graded defense and acceptability in Section 3.1, Walton emphasises that it is not feasible for reasoning agents to establish beyond doubt the validity of a scheme's *presumptions* if one is to effectively engage in epistemic or practical reasoning. The presumptive nature of the grounds used to establish a claim means that any given argument instantiating a scheme, does not in and of itself provide grounds for acceptance of the claim as having the status of true justified belief, or being the decision option that indisputably maximises a given objective. Rather, one establishes confidence in the claim sufficient to reason or act on the basis of the claim [36]. Moreover, each scheme is associated with critical questions that render explicit the presumptive nature of an argument's grounds. For example, 'Is E an expert in domain D?' and 'Is E reliable as as source ?'. A natural way to formalise reasoning with argument schemes is to regard them as defeasible inference rules and to regard critical questions as pointers to counter-arguments that may themselves instantiate schemes [37, 38]. *ASPIC+* arguments are built from such defeasible rules defined over a first order language, and premises that are first order formulae. Arguments can be *rebut* attacked on the conclusions of an argument's defeasible rules, *undermine* attacked on the argument's premises, or *undercut* attacked by challenging the applicability of a defeasible rule (via a naming mechanism for rules such that the attacking argument
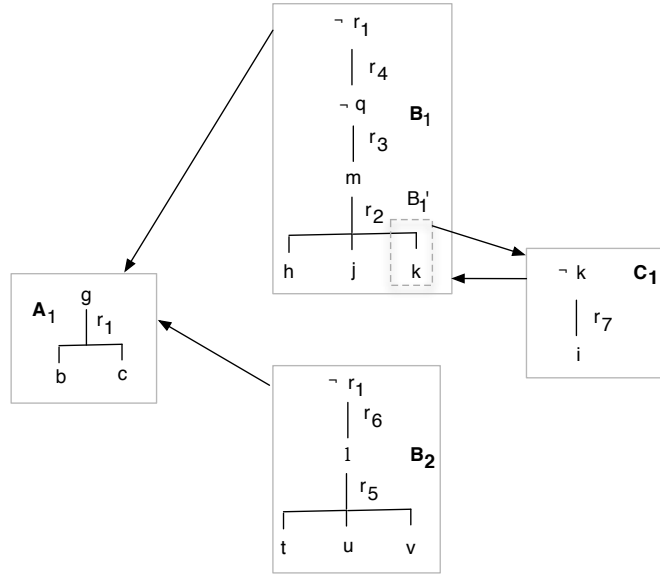
Figure 9: *ASPIC+* arguments are upside down trees whose roots are the arguments' claims and whose leaves are the arguments' premises. Notice that C_1 asymmetrically undermine attacks B_1 on its sub-argument B_1'. C_1 and B_1' also symmetrically attack (C_1 undermine attacks B_1' and B_1' rebut attacks C_1).

concludes the negation of the name of the rule in the attacked argument). For simplicity of presentation we will illustrate using rule based formulations of schemes defined over a propositional language.

Contravening the notational conventions used thus far, but in line with the standard notation in the instantiations literature, we will in this section use upper case Roman letters $A, B, C, \dots$ to denote arguments, lower case roman letters $a, b, c, \dots$ for propositions, Greek letters $\alpha, \beta, \gamma, \dots$ for variables in schemes, and notation $[a], [b], [c], \dots$ to denote subarguments consisting of a single proposition. We will use the following (abbreviated) schemes and (selected) critical questions, and instantiations of these schemes represented as propositional defeasible inference rules (whose names are shown as subscripts on the defeasible inference symbol $\Rightarrow$):

Argument from position to know. (APK):

- Source α is in a position to know about proposition γ;

- Source α asserts that γ is true (false);

- Therefore presumably γ is true (false).

APK critical questions: (APK1) Is α in a position to know about proposition γ?; (APK2) Is α an honest/trustworthy/reliable source?; (APK3) Did α assert that γ is true (false)?

33

Argument from expert opinion. (AEO):

- α is an expert in domain β;

- The domain β contains proposition γ;

- α asserts that γ is true (false);

- Therefore presumably γ is true (false).

AEO critical questions: (AEO1) how credible an expert is α?; (AEO2) Is α an expert in domain β?; (AEO3) Is α reliable as a source?; (AEO4) Did α assert that γ is true (false)?; (AEO5) Is γ consistent with what other experts assert?

6.1.2. A Detailed Example

Example 13. *Assume the premises b = 'Blair is in a position to know about whether removing Assad will achieve democracy in Syria'[25] and c = 'Blair asserts that removing Assad will achieve democracy in Syria is true', and the rule $b, c \Rightarrow_{r_1} g$ where g denotes 'removing Assad will achieve democracy in Syria'. We then have the argument A_1 in Figure 9.*

Assume furthermore the premises h = 'Chilcot is an expert in the domain of Blair's conduct in the Iraq war', j = 'the domain Blair's conduct in the Iraq war contains proposition Blair lied about weapons of mass destruction (WMD)', k = 'Chilcot asserts that Blair lied about WMD', and the rules $h, j, k \Rightarrow_{r_2} m$ (where m = 'Blair lied about WMD'), $m \Rightarrow_{r_3} \neg q$ (where q = 'Blair is an honest source'), and $\neg q \Rightarrow_{r_4} \neg r_1$. We then have the argument B_1 in Figure 9, where B_1 addresses critical question APK2 and undercuts A_1 on the inference r_1.

Intuitively, B_1's claim does not have the status of an incontrovertible belief, since its grounds do not have such status.[26] Hence, one may retain some residual confidence in A_1's claim. Graded semantics allow for this level of granularity. So, in the AF containing A_1 and B_1, A_1 and B_1 are in the 222-grounded extension.[27]

Clearly however, both our contextual (Definition 9, the natural context being $\emptyset$) and absolute (Definition 10) approaches for ranking arguments would rank B_1 above A_1.

Example 14. *Suppose an additional instance of the AEO scheme constructed from the premises t = 'The Arab League is an expert in the domain of Blair's knowledge of middle east affairs' ; u = 'the domain Blair's knowledge of middle east affairs contains the proposition Blair is an unreliable source' ; v = 'The Arab League assert that Blair is an unreliable source', and the rules $t, u, v \Rightarrow_{r_5} l$ (where l = 'Blair is an unreliable source'), $l \Rightarrow_{r_6} \neg r_1$. We then have argument B_2 in Figure 9, where B_2*

---

[25]Ex UK prime minister Tony Blair was appointed middle east envoy for peace in 2007.

[26]Commentators varyingly interpreted the Chilcot report's investigation into the Iraq war (www.iraqinquiry.org.uk/the-report) as asserting that Blair did/did not lie.

[27]All sub-arguments of A_1 and B_1, i.e., $[b]$, $[c]$, $[h]$, $[j]$, $[k]$, and the sub-arguments of B_1 that conclude m and $\neg q$, are also in the 222-grounded extension (none of these additional arguments attack or are attacked by any other arguments).
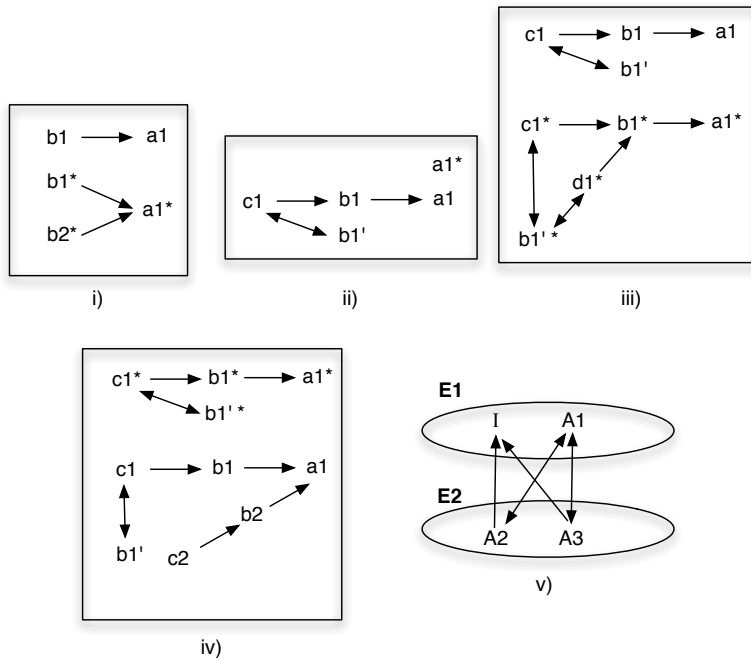
Figure 10: AFs in which $a1^* \succ^{\emptyset} a1$, illustrating the analysis of arguments in Example 14 (i), Example 15 (ii), Example 16 (iii), and Example 17 (iv). In v) we show the two Dung extensions for Example 18.

addresses critical question APK2 (by claiming unreliability rather than dishonesty) and
undercuts A_1 on the inference r_1. We now have two unattacked arguments attacking
A_1. A_1 is now further weakened, and neither $\{A_1, B_1\}$ nor $\{A_1, B_2\}$ are included in
222-preferred extensions (the threshold of allowing one unattacked attacker on A_1 and
thus maintaining co-acceptability of A_1 with one such attacker, has been exceeded).
Abstractly, this equates with the ranking $a1 \succ^S a1^*$ (according to Definition 10) in the
AF shown in Figure 10i.

In the remainder of this section we will focus on ℓmn-preferred extensions and assume S to be of such type.

Example 15. *Suppose we now have only A_1, B_1 (and their sub-arguments) and the*
argument C_1 that cites an interpretation of the report $i =$ 'Chilcot did not use the word
lie'[28] to conclude that Chilcot did not assert that Blair lied about WMD ($\neg k$), so that

---

[28]Chilcot did not use the word "lie". His report specified that it "is not questioning" Mr Blair's fixed belief, but the former Prime Minister deliberately blurred the distinction between what he believed and what he actually knew.

C_1 *undermine attacks* B_1 *(on its sub-argument* $B_1' = [k]$*), and* C_1 *undermine attacks* B_1' *and* B_1' *rebut attacks* C_1 *(Figure 9). In an AF consisting only of* A_1 *and its subarguments* $A_1' = [b]$ *and* $A_1'' = [c]$*, then for all* l, m, n *set* $\{A_1, A_1', A_1''\}$ *would be the single* lmn*-complete extension. However, in the AF that includes* $C_1 \to B_1 \to A_1$, $C_1 \leftrightarrow B1'$ *(the additional subarguments of* C_1, B_1 *and* A_1 *are not listed here as none of these are involved in attacks)* A_1 *is not in every* lmn*-complete extension (for instance,* $\{A_1, C_1\}$ *is not 112-admissible). Abstractly, this equates with* $a1^* \succ^S a1$ *in the AF shown in Figure 10ii (absolute ranking), and* $a1^* \succ^\emptyset a1$ *(contextual ranking).*

We note with interest that the higher ranking for unattacked versus reinstated arguments (illustrated by the above example) is supported by experimental findings reporting that human subjects appear to have higher confidence in claims of arguments that are unattacked, than when those arguments are subsequently attacked and then defended [39]. This suggests that our theory of graded acceptability incorporates features of human argumentation.[29]

Example 16. *Suppose the AF* Δ *that includes the arguments* $A1$, B_1 *and* C_1 *in Figure 9, and the AF* Δ' *that includes the additional argument* D_1 *that cites another interpretation of the report that concludes* $\neg k$*. Then* D_1 *also undermines attack* B_1 *on its sub-argument* B_1' *and we also now have* $D_1 \leftrightarrow B1'$*. The defense of* A_1 *against* B_1 *by* $\{C_1, D_1\}$ *is now stronger than that offered by* $\{C_1\}$ *alone. Abstractly, this equates with* $a1^* \succ^\emptyset a1$ *in the AF shown in Figure 10iii (contextual ranking). Furthermore,* $\{D_1, C_1, A_1\}$ *is a 112-preferred extension of* Δ' *but* $\{C_1, A_1\}$ *is not a 112-preferred extension of* Δ*. Abstractly, this equates with* $a1^* \succ^S a1$ *in the AF shown in Figure 10iii (absolute ranking).*

Example 17. *Suppose the AF* Δ *that includes the arguments* $A1$, B_1 *and* C_1 *in Figure 9, and the AF* Δ' *that now includes the additional argument* B_2 *in Figure 9, as well as the argument* C_2 *claiming that the Arab League are not credible experts, so that* C_2 *undercut attacks* B_2 *(on* $r5$*). That is, we have that* Δ' *includes:*

$$C_1 \to B_1 \to A_1, \, C_1 \leftrightarrow B1', \, C_2 \to B_2 \to A_1.$$

Intuitively, A_1 *is more strongly justified in* Δ *than in* Δ' *since in the latter case we have two arguments (*B_1 *and* B_2*) that continue to exert a weakening effect on* A_1 *as opposed to the one (*B_1*) in* Δ*.* A_1 *is in a 222-preferred extension of* Δ *but* A_1 *is not in a 222-preferred extension of* Δ'*. This equates with* $a1^* \succ_\Delta^S a1$ *in the AF shown in Figure 10iv.*

6.1.3. Graded Acceptability and Accrual

We now briefly illustrate, by means of a simple example, the relationship between graded acceptability and the notion of accrual. In the following section we then show how graded acceptability captures a simple form of accrual when applying graded semantics to a dialectical characterisation of non-monotonic inference.

---

[29]We do see the theory of graded acceptability also as a contribution to the long term goal of providing formal frameworks of argumentation accommodating both computational and human argumentation [40].

Example 18. *Suppose an argument in support of invading Syria, instantiating the*
scheme for practical reasoning [41]:

> *Assad is suppressing Syrians, and invading Syria will remove Assad from*
> *power, and **removing Assad will achieve democracy in Syria**, so promot-*
> *ing the value of peace.*

Now, as well as pointing to counter-arguments, critical questions can also be posed as
challenges to presumptions, shifting the burden of proof so as to provide an argument in
support of the presumption. So the critical question 'Do the consequences of the action
achieve the goal ?' can be posed to the presumption emphasised in bold in the above
argument (with the consequences being 'the removal of Assad' and the goal being
'achieve democracy in Syria'). Suppose that in response to this challenge, the above
argument is then extended to an argument I that now includes as a sub-argument, the
argument A_1 in Figure 9 whose conclusion is the bold text in the above. Now suppose
two additional arguments A_2 and A_3, each of which are instances of the AEO scheme.
A2 cites the Institute of Middle Eastern Studies at King's College London who are
experts in the domain of middle east politics and who assert that removing Assad will
not achieve democracy in Syria (i.e., A2 concludes $\neg g$). A_3 cites the United Nations
working group on middle east affairs who are also experts that assert that removing
Assad will not achieve democracy in Syria. Both A_2 and A_3 symmetrically rebut attack
A_1. Moreover, both A2 and A3 asymmetrically rebut attack I on its sub-argument A1.
We thus have two Dung admissible extensions E1 and E2 in Figure 10v.

However, A_2 and A_3 accrue in support of each other's claim ($\neg g$) and so strengthen
each other at the expense of A_1. Hence, although E1 and E2 are both subsets of Dung
preferred extensions, we have that $A2, A3 \succ_\Delta^S A1, I$ (intuitively each attack on I and
A_1 is defended by one counter-attacker, whereas each attack on A_2 and A_3 is defended
by two counter-attackers).

Observe also that C_1 and D_1 accrue in support of $\neg k$ to strengthen A_1 in Example
16. The incompatibility of accrual with Dung's standard theory has been discussed and
argued for in [26]. However, as illustrated in the above examples, our graded semantics
partially challenges this view by showing how a simple counting-based form of accrual
can be coherently accommodated within Dung's theory.

6.2. Graded Semantics and Characterisations of Non-monotonic Inference

A number of works [8, 42, 43] provide argumentation-based characterisations of
non-monotonic inference defined by Brewka's Preferred Subtheories [24]. The latter
starts with a a *totally* ordered ($\leq$) set $\mathcal{B}$ of classical wff [30] partitioned into equivalence
classes $(\mathcal{B}_1, \ldots, \mathcal{B}_n)$ (for $i = 1 \ldots n$, $\alpha, \beta \in \mathcal{B}_i$ iff $\alpha \approx \beta$) and such that:

$\forall \alpha \in \mathcal{B}_i, \forall \beta \in \mathcal{B}_j, i < j$ iff $\beta < \alpha$.

A 'preferred subtheory' ('*ps*' for short) is obtained by taking a maximal under set
inclusion consistent subset of $\mathcal{B}_1$, maximally extended with a subset of $\mathcal{B}_2$, and so on.

---

[30]Where as usual, $\alpha \approx \beta$ iff $\beta \leq \alpha$ and $\alpha \leq \beta$, and $\beta < \alpha$ iff $\beta \leq \alpha$ and $\alpha \nleq \beta$.
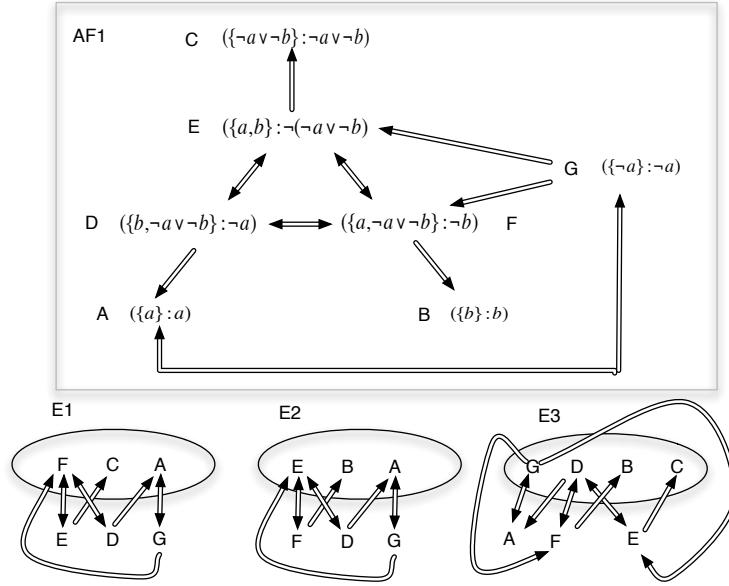
Figure 11: Arguments in $AF1$ that defeat each other are connected by double headed arrows. $E1$, $E2$ and $E3$ are Dung admissible (i.e., 111-admissible extensions). Only $E3$ is 112-admissible.

In this way, multiple *ps*s may be constructed; for example $(\mathcal{B}_1 = \{\neg a \vee \neg b\}, \mathcal{B}_2 = \{a, b\})$ yields the *ps*s $\{\neg a \vee \neg b, a\}$ and $\{\neg a \vee \neg b, b\}$.

Classical logic arguments in *ASPIC+* consist of consistent subsets of premises $\Delta \subseteq \mathcal{B}$ in support of a claim α classically entailed by Δ (and such that no proper subset of Δ entails α) via a strict inference rule $\alpha_1, \ldots, \alpha_n \rightarrow \alpha$ ($\Delta = \bigcup_{i=1}^{n} \alpha_i$) that encodes the classical entailment, and which we represent here as (Δ, α). The arguments constructed from $\mathcal{B}$ are then related by a binary defeat relation, obtained on the basis of the attacks and a strict preference relation over arguments defined by reference to the ordering on $\mathcal{B}$:

- $(\Delta, \alpha) \prec (\Delta', \alpha')$ iff $\exists \gamma \in \Delta, \forall \beta \in \Delta'$: $\gamma < \beta$;
- (Δ_1, α_1) *attacks* (Δ_2, α_2) on $(\{\beta\}, \beta)$ iff $\beta \in \Delta_2, \beta = \neg\alpha_1$ or $\alpha_1 = \neg\beta$; (Δ_1, α_1) *defeats* (Δ_2, α_2) if $(\Delta_1, \alpha_1) \not\prec (\{\beta\}, \beta)$.

[8] then shows a correspondence such that each *ps* of $\mathcal{B}$ is the set of premises in a stable extension of the AF consisting of the arguments and defeats defined by the totally ordered $\mathcal{B}$. Then α is a sceptical (credulous) *ps*-inference iff α is entailed by all (respectively at least one) *ps*, iff α is a sceptical (credulous) argumentation defined inference (i.e., α is the conclusion of an argument in all, respectively at least one, stable extension). Moreover, it has subsequently been shown in [44], that the stable extensions of any AF of classical logic arguments and defeats, as defined above by a totally
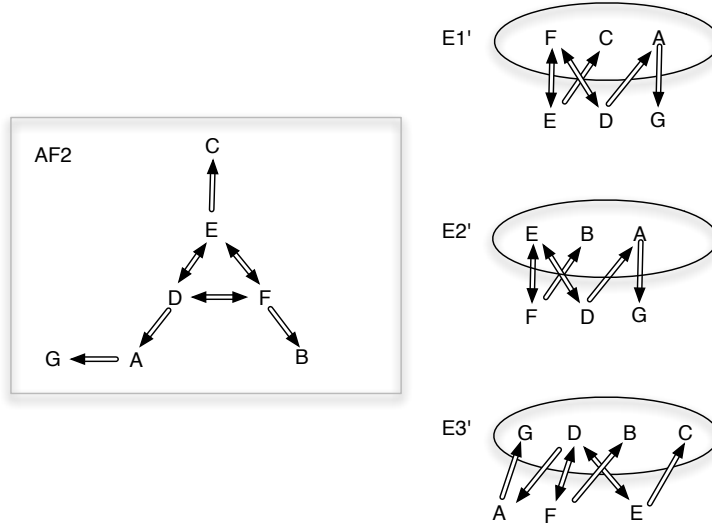
Figure 12: $AF2$ defined by the base in which $\neg a$ is ordered below a, b and $\neg a \vee \neg b$. $E1'$, $E2'$ and $E3'$ are 111-admissible extensions. None are 112-admissible.

ordered set $\mathcal{B}$ of classical logic formulae, coincide with the preferred extensions.[31] The correspondence in [8] can therefore also be stated for preferred subtheories and pre-
ferred extensions, and so in what follows, we focus on graded preferred semantics.

We can apply graded semantics to AFs that relate arguments by *defeats* rather than attacks. Firstly, assume $\mathcal{B}$ is the single equivalence class $\mathcal{B}_1 = \{a, b, \neg a \vee \neg b, \neg a\}$, gen-
erating the defeat graph $AF1$ in Figure 11 in which all attacks succeed as defeats (rep-
resented as double arrows). Note that for every argument (Δ, α) shown, the AF also
includes the argument (Δ, β) where β is any classical consequence of Δ. However, no attacks originate from these additional arguments. It therefore suffices to consider only the arguments shown in Figure 11, since these are the only arguments from which attacks originate. Hence if an extension of $AF1$ contains a (Δ, α) shown in Figure 11, then it will also contain any (Δ, β) such that $\Delta \vdash \beta$.

Now observe that $E3$ is a subset of the single 112-preferred extension that in-
cludes all arguments with consistent subsets of $\{b, \neg a \vee \neg b, \neg a\}$ (every defeat on $X \in E3$ is defended by two arguments in $E3$). However $E1$ and $E2$ are not 112-
admissible, and are not subsets of 112-preferred extensions. Indeed, we obtain that $G, D, B, C \succ_{AF1}^{\ell mn-preferred} A, E, F$ according to Definition 10. Once again we wit-
ness how graded semantics can effectively account for a form of accrual. We have two distinct arguments claiming $\neg a$ (G and D) which accrue so as to strengthen each other, and so privilege the inference $\neg a$ over a. Moreover these arguments accrue in their

---

[31]It is well known that any stable extension of an AF is preferred. Therefore [44] focuses on showing that any preferred extension is also stable.

defense of B, C and D.

Notice that explicit preferences over arguments take precedence over the implicit preference yielded by accruing arguments and the arguments they defend. Suppose that $\mathcal{B}$ consists of the two equivalence classes $\mathcal{B}_1 = \{a, b, \neg a \vee \neg b\}$ and $\mathcal{B}_2 = \{\neg a\}$ (i.e., $\neg a < a, b, \neg a \vee \neg b$) generating the defeat graph $AF2$ in Figure 12. The fact that $G \prec A$ now means that G and D no longer accrue in support of $\neg a$ and an implicit preference for $\neg a$, and no longer accrue in defense of B, C and D. Now none of $E1'$, $E2'$ and $E3'$ are subsets of 112-preferred extensions, and $\forall X, Y \in \{G, D, B, C, A, E, F\}$, $X \approx_{AF2}^{\ell mn - preferred} Y$.

7. Related Work

We discuss related work and provide a detailed comparison of our semantics based on graded acceptability with some of the approaches proposed in the literature.

7.1. Approaches to Graduality in Abstract Argumentation

The existing literature aiming at introducing some form of 'graduality' or 'ranking' in abstract argumentation can roughly be classified in two strands: those introducing a more fine-grained notion of argument status, and which are closely based on Dung's theory; those that depart from Dung's notions of defense, acceptablity and extensions under various semantics, so as to provide a unique ranking on arguments, typically through the use of techniques for value propagation on graphs.

Within the first category, [16] pushes the envelope of Dung's theory by showing how standard notions (e.g., belonging to or being attacked by the grounded extension, belonging to some or no admissible sets, and being attacked by some or no admissible sets) suffice to isolate six different statuses of arguments, without introducing an explicit notion of graduality. In *Weighted Argumentation Systems* (*WAS*) [12], weights are associated with attacks, and Dung extensions are generalised to relax the requirement for conflict-freeness, and allow for extensions whose contained attacks' summative weight does not exceed some predefined 'inconsistency budget'. The case where the weight of each attack is 1 (which we refer to as *WAS*$_1$), and in which the inconsistency budget therefore equates with the number of attacks tolerated in an extension, invites comparison with our $\ell 11$ graded semantics accommodating extensions that are not necessarily Dung conflict free (i.e, when $\ell > 1$).[32] The first thing to note is that *WAS*$_1$ considers the neutrality of an extension w.r.t. the number of contained attacks, rather than w.r.t. the number of attacks on any given contained argument. Moreover, we have argued that toleration w.r.t. the co-acceptability of attacking arguments should go hand in hand with toleration w.r.t. weaker forms of defence. However, the implications of relaxing conflict-freeness on the existence and construction of Dung's semantics in *WAS* have yet to be studied in depth ([12] only notes that the existence of a *unique* grounded extension is not guaranteed).

---

[32]We will in Section 8 suggest future research in which we generalise graded neutrality, defense and semantics to account for weights on attacks.

Amongst the graph propagation approaches, we mention the equational approaches that assign a more fine grained ranking to arguments by evaluating fixed points of functions that assign a numerical value to any given argument based on the values of its attackers. In particular, the equational approach of Gabbay and Rodrigues [45] who conjecture that their approach yields a unique solution for cyclic graphs, the compensation based semantics of Amgoud et.al. [18], which assigns the same ranking to all arguments in cycles and yield a unique solution for cyclic graphs, and the social argumentation approach of Leite and Martin (**LM**) [23] (who are concerned with propagating user votes on arguments, but also yield fine grained rankings without recourse to exogenous information).

Other graph propagation approaches that do not use an equational fixed point approach, typically account for the attack and defense paths terminating in the argument being ranked, where these paths are sequences of, respectively even and odd, numbers of attacking arguments. Besnard and Hunter (**BH**) [14] rank classical logic arguments in acyclic graphs, through a *categoriser* function – $v : \mathcal{A} \to [0, 1]$ defined as $v(a) = \frac{1}{1 + \sum_{b \to a} v(b)}$ – that assigns high values to arguments with low-valued attackers (and the maximum value to un-attacked arguments) and low values to arguments with high-valued attackers. Cayrol and Lagasquie-Schiex (**CL**) [15] then generalises use of this function to Dung AFs to develop (in their terms) a 'local approach' to valuation of arguments, and then formalise a 'global approach' that they argue gives more intuitive outcomes. Their approach requires a highly involved transformation of cyclic graphs to infinite acyclic graphs. More recently, Amgoud and Ben-Naim [17] propose two ranking-based semantics, which they call discussion-based (**AB-d**) and burden-based (**AB-b**). These semantics are also based on the processing of attack paths and their general applicability relies on conjectures concerning the processing of cyclic paths.

Finally, Matt and Toni (**MT**) [19] provide a highly original paradigm for ranking arguments, defining argument strength in terms of the value of a repeated two-person zero-sum strategic game with imperfect information.

7.2. Graded Semantics in the Landscape of Ranking-Based Semantics

Properties of rankings have been proposed by the above works, and a very informative comparison of these approaches—in particular **LM**, **BH**, **CL** (for acyclic AFs), **AB-d**, **AB-b** and **MT**—in terms of whether or not these properties are satisfied, has been provided in [20]. It is instructive to study these properties as they apply to our graded rankings, in part because such a study reveals fundamental distinctions between propagation based and Dung semantics based approaches to evaluation of arguments.

In what follows, we refer to absolute graded rankings $x \succeq^S y$ (Definition 10), which are the more naturally comparable with the other existing approaches to graduality. Unless stated otherwise, we assume S stands for any of the semantics in $\{grounded, preferred, stable\}$. Moreover, we assume the constraint $n \geq m$ and $l \geq m$. Albeit not essential for the comparison, this assumption streamlines some of the proofs in this section. We now turn to the properties studied in [20], and discuss whether our approach satisfies each of them. In doing so we will recall which approaches satisfy each property (writing **All** if satisfied by all, and **None** if satisfied by none).

7.2.1. Abstraction and Independence

Abstraction (**All**) states that rankings of arguments are preserved by isomorphisms. It should be obvious to see that graded rankings satisfy this property.

Independence (**All**) states that the ranking of a and b should be independent of any argument that is neither connected to a or b. Formally, the connected components of an AF Δ are the set of largest subgraphs of Δ, denoted $cc(\Delta)$, where two arguments are in the same component iff there is some path of attacks (ignoring the direction of attack) between them. Graded rankings satisfy independence:

Proposition 1. $\forall \Delta' = (\mathcal{A}', \rightarrow') \in cc(\Delta = (\mathcal{A}, \rightarrow))$, $x, y \in \mathcal{A}'$, and positive integers ℓ, m and n such that $\ell \geq m$ and $n \geq m$:

$$x \succeq^S_{\Delta'} y \Longrightarrow x \succeq^S_\Delta y. \tag{25}$$

Proof. To prove (25) we establish the following diagram of implications. For any integers ℓ, m and n such that $\ell \geq m$ and $n \geq m$, for any extension $S \in \{grounded, preferred, stable\}$ such that $S_{\ell mn}(\Delta) \neq \emptyset$:[33]

$$
\begin{array}{ccll}
y \in \bigcap S_{\ell mn}(\Delta') & \Longrightarrow & x \in \bigcap S_{\ell mn}(\Delta') & \text{Assumption} \\
\Updownarrow & & \Updownarrow & \boxed{\text{Claim}} \\
y \in \bigcap S_{\ell mn}(\Delta) & \Longrightarrow & x \in \bigcap S_{\ell mn}(\Delta) & \text{Conclusion}
\end{array}
$$

for $x, y \in \Delta'$. We need to establish the equivalences denoted $\boxed{\text{Claim}}$ in the above diagram, for each of the semantics. Firstly note that, clearly, since $\Delta' \in cc(\Delta)$:[34]

$$\forall x \in \Delta' : \overline{x}_{\Delta'} = \overline{x}_\Delta \text{ AND } \overline{\overline{x}}_{\Delta'} = \overline{\overline{x}}_\Delta \tag{26}$$

$\boxed{S = grounded}$ It suffices to show that $\forall lmn$ satisfying the constraint we have that $\text{grn}_{\ell mn}(\Delta') = \text{grn}_{\ell mn}(\Delta) \cap \mathcal{A}'$. We proceed by induction.[35] By Corollary 1 $\text{grn}_{lmn}(\Delta) = \bigcup_{0 \leq k < \omega} \text{d}_{\underset{n}{m}}^k(\emptyset)$. For the base case, clearly $\text{d}_{\underset{n}{m}}^{\Delta'}(\emptyset) = \text{d}_{\underset{n}{m}}^\Delta(\emptyset) \cap \mathcal{A}'$. Assume then that for $i > 1$ $(\text{d}_{\underset{n}{m}}^{\Delta'})^i(\emptyset) = (\text{d}_{\underset{n}{m}}^\Delta)^i(\emptyset) \cap \mathcal{A}'$ (IH). By (26) and IH it immediately follows that the same claim holds for $i + 1$ and therefore, in the limit, $\text{grn}_{\ell mn}(\Delta') = \text{grn}_{\ell mn}(\Delta) \cap \mathcal{A}'$ as desired.

$\boxed{S = preferred}$ First of all, one can straightforwardly show that $\forall E$, if $E \in \text{prf}_{\ell mn}(\Delta)$ then $E \cap \mathcal{A}' \in \text{prf}_{lmn}(\Delta')$. So for any lmn-preferred extension in Δ there is a corresponding lmn-preferred extension in Δ'. To establish the claim it then suffices to prove the converse, that is:

$$\forall E' \in \text{prf}_{lmn}(\Delta'), \exists E \in \text{prf}_{\ell mn}(\Delta) \text{ S.T. } E' = E \cap \mathcal{A}'. \tag{27}$$

---

[33]This condition is obviously relevant only for $S = stable$.

[34]Recall notation introduced in Definition 1.

[35]To illustrate the argument we use here standard induction, which to be precise works only for the case in which Δ is finitary. For the general case one needs to generalize the argument in the obvious way to transfinite induction. Cf. Remark 1.

To prove (27) an inductive argument similar to the above one for $S = grounded$ can be used by exploiting Fact 8, for each ℓmn-preferred extension. The claim then follows again from (26).

$\boxed{S = stable}$. Assume $\mathtt{stb}_{\ell mn}(\Delta) \neq \emptyset$. It follows that $\mathtt{stb}_{\ell mn}(\Delta') \neq \emptyset$. Then the same argument used for the case of preferred extensions applies to prove $\boxed{\text{Claim}}$. If instead, $\mathtt{stb}_{\ell mn}(\Delta) = \emptyset$ then it trivially holds that $x \in \bigcap \mathtt{stb}_{\ell mn}(\Delta) \Leftrightarrow y \in \bigcap \mathtt{stb}_{\ell mn}(\Delta)$, no matter whether $y \in \bigcap \mathtt{stb}_{\ell mn}(\Delta') \Rightarrow x \in \bigcap \mathtt{stb}_{\ell mn}(\Delta')$. This suffices to establish (25) for $S = stable$. $\qquad\square$

An inspection of the proof of Proposition 1 should show, however, that the claim can be strengthened to $\forall \Delta' \in cc(\Delta)$: $x \succ_{\Delta'}^S y \Rightarrow x \succ_{\Delta}^S y$, only for $S \in \{grounded, preferred\}$. This stronger formulation fails for $S = stable$ because of the well known non-existence of Dung stable extensions due to arguments in an odd cycle 'contaminating' unrelated arguments. For example, in

$$\Delta = \langle \{a, b, c\}, \{a \rightarrow a, b \rightarrow c\} \rangle \text{ and } \Delta' \in cc(\Delta) = \langle \{b, c\}, \{b \rightarrow c\} \rangle$$

we have that $b \succ_{\Delta'}^{stable} c$ since b, but not c, is in the single 111-stable—i.e., Dung stable—extension. However for Δ, there are no ℓ, m and n such that b, but not c, is justified under ℓmn-stable semantics.

So with respect to the properties of abstraction and independence, graded rankings behave in line with most existing approaches to graduality, with some important differences when the underlying semantics is assumed to be the stable one.

7.2.2. *Void Precedence and Self Contradiction*

The above example also illustrates why **Void Precedence** (**All**)—any non-attacked argument is ranked strictly higher than any attacked argument—is also not satisfied by graded rankings under stable semantics, as $b \approx_{\Delta}^{stable} c$. However, the situation is more positive under grounded and preferred semantics, for which an—arguably—natural weakening of the property holds.

We need some auxiliary notation. Define the variant of the graded defense function (Definition 5) that requires an infinity of defenders for any attack: $\mathtt{d}_{1_\infty} = \{x \in \mathcal{A} \mid \forall y \text{ s.t. } y \rightarrow x, \exists^\infty z \in X \text{ s.t. } z \rightarrow y\}$. So this function outputs, for any set, the set of arguments whose attackers are counter-attacked by an infinity of arguments in that set. Consider now the smallest fixpoint $\mathsf{lfp}.\mathtt{d}_1$ of such a function. We can refer to this as the 11∞-grounded extension. Observe right away that if the underlying framework is finitary, then $\mathsf{lfp}.\mathtt{d}_{1_\infty}$ is equal to the set of unattacked arguments in that framework. We have the following result:

Proposition 2. *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF and let $x \in \mathcal{A}$ be s.t. $x \in \mathsf{lfp}.\mathtt{d}_{1_\infty}^\Delta$. Then $\forall y \in \mathcal{A} \backslash \mathsf{lfp}.\mathtt{d}_{1_\infty}^\Delta$ s.t. $\overline{y} \neq \emptyset$: $x \succ_{\Delta}^S y$, for $S \in \{grounded, preferred\}$.*

Proof. Let $y \in \mathcal{A} \backslash \mathsf{lfp}.\mathtt{d}_{1_\infty}^\Delta$. Then $\overline{y} \neq \emptyset$ and $\overline{\overline{y}}$ is finite. Then let $k = \max \{|\overline{z}| \mid z \in \overline{y}\}$. It follows that there exists no $11k + 1$-admissible (and hence grounded or preferred) extension which contains y. Hence, if $x \in \mathsf{lfp}.\mathtt{d}_{1_\infty}^\Delta$ then $x \succ_{\Delta}^S y$. $\qquad\square$

A direct consequence of the proposition is that, if Δ is finitary, the claim simplifies to the standard formulation of void preference, that is: $\forall x \in \mathcal{A}$ s.t. $\overline{x} = \emptyset$, then $\forall y \in \mathcal{A}$ s.t. $\overline{y} \neq \emptyset$: $x \succ_{\Delta}^{S} y$, $S \in \{grounded, preferred\}$.

Finally, and related to the above discussion, the property of **Self Contradiction** (**MT**), which states that a self-attacking argument is ranked strictly lower than any non self-attacking argument, is also not satisfied by our approach. Consider $\Delta = \langle \{a, a1, a2, b\}, \{a \rightarrow a, a1 \rightarrow b, a2 \rightarrow b\} \rangle$, where a is in all 222-admissible (and hence 222-preferred) extensions, whereas b is not in any 222-admissible extension, and so $b \not\succeq_{\Delta}^{preferred} a$.

So with respect to the properties of void precedence and self contradiction, graded rankings behave essentially in line with most existing approaches to graduality. Some differences arise with respect to void precedence, when the underlying semantics is assumed to be the stable one and when frameworks are not finitary, in which case graded rankings cannot distinguish between unattacked arguments and arguments whose attackers are recursively counter-attacked by an infinity of defenders ($\mathsf{lfp.d}_{1}{}_{\infty}$).

7.2.3. Further Postulates

Cardinality Precedence (**AB-d**,**AB-b**) states that if the number of attacks on b is strictly greater than the number of attacks on a, then a is ranked strictly higher than b. Like **LM**, **BH**, **CL** and **MT**, our approach does not satisfy this postulate. This is because our ranking also accounts for the number of defenders, as witnessed by the incomparability of $a3$ and $a4$ in Figure 4, and the fact that $a3 \not\succeq_{AF}^{grounded} a4$ and $a4 \not\succeq_{AF}^{grounded} a3$ (recall Example 12). It is worth remarking however that, as discussed in Section 3.4.1, our ranking could be further refined to a lexicographic ordering giving precedence to the minimization of attackers, that would then satisfy the cardinality precedence postulate.

Quality Precedence (**None**) states that if there is an attacker of b that is ranked strictly higher than all attackers of a, then a is ranked strictly higher than b. The principle is not satisfied by our approach. We leave it to the reader to verify that given $e \rightarrow c \rightarrow b$, $e1 \rightarrow d1 \rightarrow a$, $e2 \rightarrow d1$, $e3 \rightarrow d2 \rightarrow a$, $e4 \rightarrow d2$, then although each of $d1$ and $d2$ are ranked lower than c, both a and b are incomparable.

Strict Counter-transitivity (**LM**, **BH**, **AB-d**,**AB-b**) states that if the attackers of b are strictly more than those of a (i.e., cardinality precedence), or the number of attackers of each are the same, but at least one attacker of b is ranked strictly higher than at least one attacker of a, and not vice versa, then a is ranked strictly higher than b. Like **CL** and **MT**, graded rankings do not satisfy this property. Consider the AF and ranking of arguments in Figure 13i), where a and b both have one attacker (respectively d and c). Then $c \succ_{AF}^{S} d$, but $a \not\succ_{AF}^{S} b$. In fact, $a \approx_{AF}^{S} b$. With respect to this, it is worth observing the following. By virtue of our generalisation of Dung semantics, analysis of any individual argument under graded semantics is inherently bound up in the analysis of sets of (co-acceptable) arguments (recall the discussion in Section 3.1). Now, observe that the strongest defense needed to accept the defender $e1$ of a, and that then suffices to defend a, is when $m = 2$ (since $e1$ is not defended). But then any such standard of defense also accepts b (since the attack by c need not then be defended).
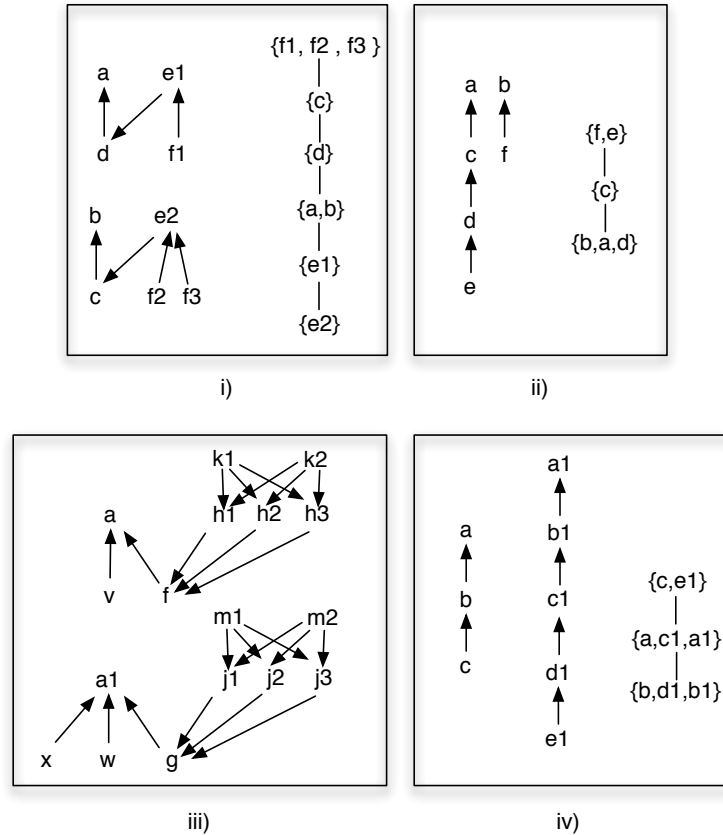
Figure 13: Frameworks and graded rankings referenced in the discussion of postulates for ranking based semantics reviewed in [20].

Defense Precedence (**LM**, **BH**, **AB-d**,**AB-b**) states that for two arguments with the same number of attackers, a defended argument is ranked strictly higher than a non-defended argument. Again, like **CL** and **MT**, this property is not satisfied by graded rankings, as illustrated in Figure 13ii), where the defended a and undefended b have the same ranking. We again see that the analysis of any individual argument is inherently bound up in the analysis of sets of (co-acceptable) arguments. In this example, the strongest defense needed to accept d and that suffices to accept a, is when $m = 2$ (since d is undefended). But then this suffices to accept b.

Cayrol and Lagasquie-Schiex [15] propose a number of properties that relate to the addition and extension of attack/defense paths.[36] Improving the ranking of an argu-

---

[36]These properties, referencing addition of and extension (increase in) paths, assume that the arguments

ment by **Strict Addition of a Defense Path**[37] (**None**) is clearly not satisfied by graded rankings. Indeed, the reverse is the case: a is ranked higher when un-attacked than when attacked and then defended (as discussed in Example 15). Restriction of this property to the case where a defense path is added only to an argument that is already attacked, is satisfied only by **CL**, and is not satisfied by graded rankings.

Satisfaction of **Addition of an Attack Path** (**All**) means that the ranking of an argument a is degraded when adding an attack path terminating in a, and equates with a being ranked strictly higher than $a1$ in the AF $\langle c \to b \to a, c1 \to b1 \to a1, x_n \to \ldots \to x1 \rangle$, where $x_n \to \ldots \to x_1$ is a path of attacks s.t. $x_1 = a_1$ and $n \in 2\mathbb{N}$. This property is in general not satisfied by graded rankings. Consider the example Δ in Figure 13iii). The strongest iterated defense from $\emptyset$ (and hence the 'strongest' lmn-grounded extension) that includes a is when $l = m = n = 3$. This defense also includes $a1$. Observe that: $\mathsf{d}_{\underset{3}{3}}(\emptyset) = E_0 = \{k1, k2, h1, h2, h3, a, v, m1, m2, j1, j2, j3, x, w\}$; $\mathsf{d}_{\underset{3}{3}}(E_0)$ $= E_1 = E_0 \cup \{a1\}$; $\mathsf{d}_{\underset{3}{3}}(E_1) = E_1$ is the 333-grounded extension. Indeed, there is no lmn-grounded extension that contains a and not $a1$. Intuitively, if $m = 2$ ($n \geq 2$), then the defense of a requires at least two of $h1$, $h2$ and $h3$ (since v is unattacked and so f must be defended against by $n \geq 2$ arguments), but then a defense of any h_i requires that $m = 3$ (since each h_i is attacked by two unattacked arguments). Hence inclusion of a in an lmn grounded extension requires that $m \geq 3$. But then any such defense allows an additional attack path terminating in $a1$, while accommodating $a1$ under any such standard of defense. Put briefly, and recalling our discussion of strict countertransitivity and defense precedence, it is the standard of defense met by $h1$, $h2$ or $h3$ that determines the strongest defense accommodating a, and this standard of defense allows for an additional attack path on $a1$.

The two properties, **Increase of an Attack Path** (**All** except **MT**) and **Increase of a Defense Path** (**All** except **MT**) are distinctive of the propagation based approaches whereby the ranking of arguments is propagated down chains of attacks. Neither is satisfied by graded rankings. The former equates with $b1$ being ranked strictly higher than b, and the latter equates with $a1$ being ranked strictly lower than a, in the AF in Figure 13iv). We have that $b \approx^S_{AF} b1$ and $a \approx^S_{AF} a1$. Once again, the strongest standard of defense needed to accept $a1$ is that needed to defend $c1$, which in turn has the same ranking as a.

To the above discussed properties, [20] also proposes that all arguments can be compared according to their ranking, which is clearly not satisfied by graded rankings (cf. discussion in Section 3.4.1), and that all non-attacked arguments have the same ranking, which is obviously satisfied by our approach.

7.2.4. Discussion

We briefly comment on the above analysis and its positioning of our proposal within the growing field of ranking-based semantics. Like all existing approaches, our rankings satisfy key properties such as abstraction and independence, as well as void prece-

---

in the additional (extension to the) path are disjoint from the existing arguments.

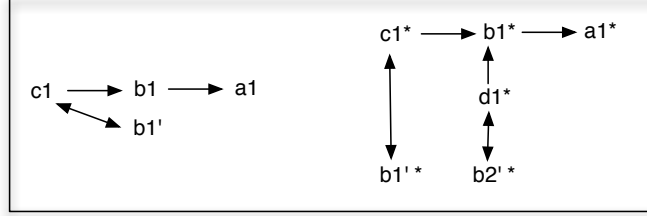[37]We use the term 'path' rather than the term 'branch' used in [20].

Figure 14: AF in which $a1^* \approx^{\emptyset} a1$.

dence, albeit with some interesting caveats concerning graded stable extensions and
the finitariness condition. With respect to other postulates, rankings based on graded
acceptability tend to behave differently from approaches based on the propagation idea
(and appear to behave more closely to **MT** which is also underpinned by intuitions
different from propagation), for reasons inherent to the way graded semantics gener-
alise standard Dung acceptability. To recap, we recall the discussion in Section 3.1
where we emphasise that our graded theory of acceptability aims at capturing a notion
of graduality while at the same time retaining the Dungian focus on *sets of* arguments,
rather than individual arguments, as the units of analysis. Hence, as our above analysis
repeatedly shows, it is the defense of co-acceptable defenders of an argument x that de-
termines the strongest defense needed to accept (and hence determine the ranking of)
x. We further illustrate this point with a variation of Section 6.1 's examples of graded
semantics applied to frameworks instantiated by schemes and critical questions.

Example 19. *Suppose the AF Δ that includes the arguments A_1, B_1 and C_1 in Figure
9, and the AF Δ' that includes the additional argument D_1 that concludes $\neg h$ (i.e.,
an argument addressing* AEO*'s critical question AEO2, by claiming that Chilcott is
not an expert in the domain of Blair's conduct in the Iraq war[38]). Notice that unlike
Example 16, D_1 undermine attack B_1 on a different sub-argument $B_2' = [h]$ to that
(i.e., B_1') undermine attacked by C_1. We now also have $D_1 \leftrightarrow B_2'$. Now, A_1 is not then
strengthened upon inclusion of D_1. Abstractly, this equates with the ranking $a1^* \approx^S
a1$ in the AF in Figure 14. Intuitively, the strongest defense needed to accept a defender
(either $c1^*$ or $d1^*$) of $a1^*$, and that then suffices to defend $a1^*$, is when $m = 1$ and $n
= 1$. But then this is also the strongest defense needed to accept $a1$. Contrast this
with Example 16 in which C_1 and D_1 accrue in support of the same conclusion $\neg k$ so
that both undermine attack B_1 on B_1'. Recall that this equates with $a1^* \succ^S a1$ in the
AF in Figure 10iii), in which the strongest defense needed to accept a defender (either
$c1^*$ or $d1^*$) of $a1^*$ that then suffices to defend $a1^*$, is when $m = 1$ and $n = 2$. This
standard of defense does not suffice to accept $a1$ in this example. Finally, notice that in
the contextual approach to rankings, we do have that $a1^* \succ^{\{c_1, c_1^*, d_1^*\}} a1$ in the AF in*

---

[38]For example, because Chilcott was not given access to the relevant documents reporting on Blair's
deliberations.

Figure 14. That is, if we commit to $\{c_1, c_1^*, d_1^*\}$ (the context), a_1^* is then ranked higher than $a1$.

All in all, our analysis shows that rankings based on graded acceptability yield an original combination of postulates with respect to existing approaches, and show how a graded generalisation of Dung semantics can contribute a rich perspective to the ranking-based semantics programme.

Finally, we observe that in contrast to the above described graph propagation approaches, we have (in Section 6 and in the above example) studied our notions of graduality and rankings as they apply to instantiated argumentation. In particular, we have studied well-known instantiations of Dung's theory yielded by the use of schemes and critical questions, and classical logic instantiations that yield dialectical characterisations of non-monotonic inference. We believe that such a study is an important complement to the use of postulates in evaluating the intuitions captured by proposals for graduality and rankings. It is with this aim in mind that Section 6 seeks to substantiate the intuitions captured by our graded generalisation of Dung's theory.

8. Conclusions and Future Work

Summary. This paper has presented three contributions. First, we have generalised Dung's standard semantics by providing graded variants of the classic extensions studied in abstract argumentation, and studied their fixpoint-theoretic behavior. In doing so, the paper also provided a comprehensive exposition of the fixpoint theory of standard Dung's semantics (Section 2) which, to the best of our knowledge, was not yet available in the literature.

The resulting graded semantics for abstract argumentation has then been shown to enable a simple way of ranking arguments according to how strongly they are justified under different graded semantics. We showed how this enables an arguably natural way of arbitrating, in abstract argumentation, among justified arguments that are credulously justified under the standard Dung semantics, without recourse to exogenous information. Such rankings have then also been applied to instantiated Dung frameworks, specifically via *ASPIC+* formalisations of 'human orientated' argumentation encoded through the use of argument schemes and critical questions, and dialectical characterisations of non-monotonic inference defined by Brewka's Preferred Subtheories. In so doing, we have also shown how the graded semantics accounts for the mutual strengthening of arguments with the same conclusion (i.e., the accrual of arguments) within the Dung paradigm.

Finally, the novel rankings have been thoroughly compared with existing approaches to ranking-based semantics, allowing us to highlight the similarities and core differences between rankings based on graded semantics and other influential approaches.

Future work. The paper has aimed at establishing foundations for a graded generalisation of Dung's argumentation theory. Based on our results, many natural avenues for future research present themselves. We confine ourselves to mentioning those that in our opinion are the most promising.

We observed in Section 3.4.1 that the rankings enabled by our graded semantics are, without further assumptions, partial. The ensuing theory of graded rankings is therefore less committal than existing proposals in the ranking semantics literature, and could therefore be extended to accommodate further pre-theoretical intuitions. For example, some application domains may warrant distinguishing amongst arguments justified under the standard Dung semantics, by assigning a higher ranking to those that have a higher number of attacks. For example, consider that scientific theories establish their credibility to the extent that they successfully defend themselves against arguments attempting to refute (attack) them. Exploring directions in which graded rankings could be extended to capture this alternative intuition in a principled manner, is clearly a natural direction of research.

Dung's original semantics were born out of an attempt to systematize patterns underpinning the semantics of all the main non-monotonic logics. Section 6.2's application of graded semantics to dialectical characterisations of Preferred Subtheories strongly suggests that graded semantics could offer a similar tool for systematizing paraconsistent non-monotonic logics. By relaxing the conflict-freeness and self-defence requirements, paraconsistent non-monotonic inference relations could be defined akin to the way in which non-monotonic inferences are defined by Dung's argumentation theory. As graded semantics enable a 'controlled' relaxation of conflict-freeness, these inference relations should be able to keep the explosivity of the ensuing logic at bay. The development of these ideas can lead, we argue, to novel interesting insights at the interface of paraconsistent logic and argumentation theory. This in turn may motivate the study of rationality postulates for argumentation [46] that are reformulated to accommodate paraconsistent inference relations.

Notice also, that in Section 6.2 we applied graded semantics to defeat graphs that employ exogenous preferences to decide which attacks succeed as defeats, and highlighted how these explicit preferences take precedence over the implicit preferences yielded by our rankings. The interaction between exogenous and endogenously defined preferences is also an obvious direction for future research.

The theory of graded argumentation as developed in this paper has focused on extracting graduality from information endogenous to standard argumentation frameworks, essentially by counting the attackers of arguments. This intuition can however be extended along lines explored by weighted argumentation frameworks [12]. We briefly sketch how this could be done, assuming now each attack is assigned a value in $\mathbb{R}^+$ by a weighting function w. Recall Definition 5 and consider the following alternative definition of graded defense:

$$\mathtt{d}_n^m(X) = \big\{ x \in \mathcal{A} \mid \rho(\big\{ (y,x) \in \overline{x} \mid \rho(y^- \cap X) < n \big\}) < m \big\} \tag{28}$$

where, for $R \subseteq \to$, $\rho(R) = \sum_{(x,y) \in R} w(x,y)$. Clearly the definition of graded defense given in Definition 5 is the special case of (28) where w assigns the weight of 1 to each attack. This generalization suggests that the logic behind graded semantics can be leveraged beyond the standard case explored here, and therefore also applied to existing proposals based on weighted argumentation (e.g., [12]).

In Section 7 we discussed differences between rankings defined by graded semantics and propagation-based approaches. Although, as we have shown, the two ap-

proaches build on different underpinning intuitions, it should be stressed that they remain compatible. In fact an interesting direction for future research would be to explore how features of the latter might be integrated with our graded approach. For example, recalling the example illustrating that strict counter-transitivity is not satis-
1350 fied, one might note that since b's attacker c is ranked above a's attacker d by the graded ranking in Figure 13i, then the ranking could be refined to favour a over b. Similarly, defense precedence could be enforced by noting that since b's attacker f is ranked strictly higher than a's attacker c (according to the graded ranking in Figure 13ii, then the ranking can be refined to rank b strictly below a. By the same reasoning,
1355 the ranking in Figure 13iii could be refined to rank $b1$ above b and so a above $a1$ (so enforcing satisfaction of 'increase of an attack path' and 'increase of a defense path'), and the ranking $a1^* \approx^S a1$ in Example 19 could be refined to obtain $a1^* \succ_\Delta^S a1$ by noting that $b1 \succ_\Delta^S b1^*$.

Finally, Dung's standard semantics lend themselves to natural dialectical interpre-
1360 tations via argument games (cf. [47] for an overview). Given that graded semantics retain many of the key fixpoint-theoretic properties of standard Dung's semantics, we believe that standard argument games—specifically the game for the grounded extension and for the credulous preferred semantics—could be appropriately modified to obtain game-theoretic characterizations of (at least some of) our semantics.[39] These types
1365 of games would shed novel light on the under-investigated link between dialectical approaches to argumentation and the younger literature on ranking-based semantics.

Appendix. Proofs of Section 2

Proof of Fact 2. $\boxed{(1)}$ We prove the two directions. $\boxed{\Leftarrow}$ The claim follows from the monotonicity of d_Δ. For any X, $\mathrm{d}_\Delta(X) \subseteq \mathrm{d}_\Delta(\bigcup_{X \in D} X)$, hence $\bigcup_{X \in D} \mathrm{d}_\Delta(X) \subseteq$
1370 $\mathrm{d}_\Delta(\bigcup_{X \in D} X)$. $\boxed{\Rightarrow}$ Assume $x \in \mathrm{d}_\Delta(\bigcup_{X \in D} X)$. By finitariness, and since D is upward directed, $\exists X \in D$ s.t. if $y \to x$, $\exists z \in X$ s.t. $z \to y$ Hence $x \in \bigcup_{X \in D} \mathrm{d}_\Delta(X)$.
$\boxed{(2)}$ We prove the two directions. $\boxed{\Rightarrow}$ The claim follows again from the monotonicity of d_Δ. For any X, $\mathrm{d}_\Delta(X) \supseteq \mathrm{d}_\Delta(\bigcap_{X \in D} X)$, hence $\bigcap_{X \in D} \mathrm{d}_\Delta(X) \supseteq \mathrm{d}_\Delta(\bigcap_{X \in D} X)$.
$\boxed{\Leftarrow}$ Assume $x \in \bigcap_{X \in D} \mathrm{d}_\Delta(X)$. By finitariness there exists a finite set $X \in D$
1375 defending x. By the assumption X is contained in $\bigcap_{X \in D} X$, for otherwise there would exist a $Y \in D$ such that $X \supset Y$ and $x \notin \mathrm{d}_\Delta(Y)$, against the assumption. We conclude that $x \in \mathrm{d}_\Delta \left(\bigcap_{X \in D} X \right)$.

$\square$

Proof of Lemma 1. $\boxed{X \subseteq \mathrm{d}_\Delta^n(X)}$ holds by the assumption that X is admissible and
1380 by the monotonicity of d (Fact 1). $\boxed{\mathrm{d}_\Delta^n(X) \subseteq \mathrm{n}_\Delta(\mathrm{d}_\Delta^n(X))}$ is proven by induction over n. The base case holds by assumption as $\mathrm{d}_\Delta^0(X) = X$ is admissible. For the induction step assume (IH) that $\mathrm{d}_\Delta^n(X) \subseteq \mathrm{n}_\Delta(\mathrm{d}_\Delta^n(X))$. We show that $\mathrm{d}_\Delta^{n+1}(X) \subseteq \mathrm{n}_\Delta(\mathrm{d}_\Delta^{n+1}(X))$. Suppose towards a contradiction that is not the case. Then there exists

---

[39]Games for all semantics could be obtained via a detour through suitable logic games. Cf. [48].

$x, y \in d_\Delta^{n+1}(X) = d_\Delta(d_\Delta^n(X))$ such that $x \leftarrow y$. By the definition of d, there exists $z \in d_\Delta^n(X)$ such that $y \leftarrow z$. But as $y \in d_\Delta^{n+1}(X)$ there exists also $w \in d_\Delta^n(X)$ such that $z \leftarrow w$. From this we conclude that $d_\Delta^n(X)$ is not conflict free, against IH. $\boxed{n_\Delta(d_\Delta^n(X)) \subseteq n_\Delta(X)}$ follows from the first claim by the antitonicity of n (Fact 1). $\qquad\square$

Proof of Lemma 2. $\boxed{(3)}$ $\boxed{\text{First}}$, we prove that $\bigcup_{0 \leq n < \omega} d_\Delta^n(X)$ is a fixpoint by the following series of equations:

$$d_\Delta\left(\bigcup_{0 \leq n < \omega} d_\Delta^n(X)\right) = \bigcup_{0 \leq n < \omega} d_\Delta(d_\Delta^n(X))$$
$$= \bigcup_{0 \leq n < \omega} d_\Delta^n(X)$$

where the first equation holds by the continuity of d_Δ (Fact 2) and the second by the fact that, since X is admissible, $X = d_\Delta^0(X) \subseteq d_\Delta^n(X)$ for any $0 \leq n < \omega$. $\boxed{\text{Second}}$, we prove that $\bigcup_{0 \leq n < \omega} d_\Delta^n(X)$ is indeed the least fixpoint containing X. Suppose, towards a contradiction that there exists Y s.t.: $X \subset Y = d_\Delta(Y) \subset \bigcup_{0 \leq n < \omega} d_\Delta^n(X)$. It follows that $X \subset Y = d_\Delta(Y) \subset d_\Delta^n(X)$ for some $0 \leq n < \omega$. But, by monotonicity, we have that $d_\Delta^n(X) \subseteq d_\Delta^n(Y)$. Contradiction.

$\boxed{(4)}$ The proof is similar to the previous case, but involves a few extra subtleties. $\boxed{\text{First}}$, we prove that $\bigcap_{0 \leq n < \omega} d_\Delta^n(n_\Delta(X))$ is a fixpoint, through the series of equations

$$d_\Delta\left(\bigcap_{0 \leq n < \omega} d_\Delta^n(n_\Delta(X))\right) = \bigcap_{0 \leq n < \omega} d_\Delta(d_\Delta^n(n_\Delta(X)))$$
$$= \bigcap_{0 \leq n < \omega} d_\Delta^n(n_\Delta(X))$$

which hold by Fact 2 and by the fact that $n_\Delta(X) = d_\Delta^0(n_\Delta(X)) \supseteq d_\Delta^n(n_\Delta(X))$ for any $0 \leq n < \omega$. The latter property holds because X is assumed to be admissible, and hence $X \subseteq d_{\Delta(X)}$. By the antitonicity of n_Δ (Fact 1) we therefore have that $n_\Delta(d_\Delta(X)) \subseteq n_\Delta(X)$ and, since $d_\Delta = n_\Delta \circ n_\Delta$ (Fact 1 again), it follows that $d_\Delta(n_\Delta(X)) \subseteq n_\Delta(X)$ as desired. $\boxed{\text{Second}}$, it remains to be proven that $\bigcap_{0 \leq n < \omega} d_\Delta^n(n_\Delta(X))$ is indeed the largest fixpoint of d_Δ contained in $n_\Delta(X)$. Like in the previous case we proceed towards a contradiction. Suppose there exists Y s.t.: $\bigcap_{0 \leq n < \omega} d_\Delta^n(n_\Delta(X)) \subset Y = d_\Delta(Y) \subseteq n_\Delta(X)$. There must therefore exist an integer k such that, as a consequence of Lemma 1, $X \subseteq d_\Delta^k(X) \subseteq n_\Delta(d_\Delta^k(X)) \subset Y$. By the antitonicity of n_Δ and the fact that $d = n \circ n$ (Fact 4), and since Y is taken to be a fixpoint of d_Δ, it follows that $n_\Delta(Y) = n_\Delta(d_\Delta(Y)) \subset d_\Delta(d_\Delta^k(X))$. So $n_\Delta(Y)$ is also a fixpoint of d_Δ, it contains X and it is included in $\bigcup_{0 \leq n < \omega} d_\Delta^n(X)$, which, by the previous claim, is the smallest fixpoint of d_Δ containing X. Contradiction. $\qquad\square$

Proof of Theorem 1. By Lemma 2, $\mathsf{lfp}_X.d_\Delta = \bigcup_{0 \leq n < \omega} d_\Delta^n(X)$. So $\bigcup_{0 \leq n < \omega} d_\Delta^n(X)$ denotes a fixpoint of d_Δ, and more specifically the smallest such fixpoint that contains

the admissible set X. By Lemma 1 we also know that such set is conflict-free. We therefore conclude that $\bigcup_{0 \leq n < \omega} \mathrm{d}_\Delta^n(X)$ is a conflict-free fixpoint of d_Δ, that is, a complete extension, and that this complete extension is the smallest such set containing X, as claimed. □

References

[1] D. Grossi, S. Modgil, On the graded acceptability of arguments, in: Proceedings of IJCAI15, AAAI Press, 2015, pp. 868–874.

[2] T. J. M. Bench-Capon, P. E. Dunne, Argumentation in artificial intelligence, Artificial Intelligence 171 (2007) 10–15.

[3] I. Rahwan, G. Simari (Eds.), Argumentation in AI, Springer-Verlag, 2009.

[4] P. M. Dung, On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games, Artificial Intelligence 77 (2) (1995) 321–358.

[5] P. Baroni, M. Caminada, M. Giacomin, An introduction to argumentation semantics, The Knowledge Engineering Review 26 (4) (2011) 365–410.

[6] L. Amgoud, C. Cayrol, A reasoning model based on the production of acceptable arguments, Annals of Mathematics and Artificial Intelligence 34 (1-3) (2002) 197–215.

[7] S. Modgil, Reasoning about preferences in argumentation frameworks, Artificial Intelligence 173 (9-10) (2009) 901–934.

[8] S. Modgil, H. Prakken, A general account of argumentation with preferences, Artificial Intelligence 195 (0) (2013) 361 – 397.

[9] T. J. M. Bench-Capon, Persuasion in practical argument using value-based argumentation frameworks, Journal of Logic and Computation 13 (3) (2003) 429–448.

[10] A. Hunter, M.Thimm, Probabilistic reasoning with abstract argumentation frameworks, Journal of Artificial Intelligence Research 59 (2017) 565–611.

[11] L. Hengfei, N. Oren, T. Norman, Probabilistic argumentation frameworks, in: Theories and Applications of Formal Argumentation: First International Workshop, TAFA 2011, 2012, pp. 1–16.

[12] P. Dunne, T.Hunter, P. McBurney, S. Parsons, M.Wooldridge, Weighed argument systems. basic definitions, algorithms and complexity results, Artificial Intelligence 175 (2) (2011) 457–86.

[13] J. Janssen, M. D. Cock, D. Vermeir, Fuzzy argumentation frameworks, in: Proceedings of the 12th Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2008, p. 513?520.

[14] P. Besnard, A. Hunter, A logic-based theory of deductive arguments, Artificial Intelligence 128 (1–2) (2001) 203–235.

[15] C. Cayrol, M.-C. Lagasquie-Schiex, Graduality in argumentation, Journal of Artificial Intelligence Research 23 (2005) 245–297.

[16] Y. Wu, M. Caminada, A labelling-based justification status of arguments, Studies in Logic 3 (4) (2010) 12–29.

[17] L. Amgoud, J. Ben-Naim, Ranking-based semantics for argumentation frameworks, in: Scalable Uncertainty Management, Springer, 2013, pp. 134–147.

[18] L. Amgoud, J. Ben-Naim, D. Doder, S. Vesic, Ranking arguments with compensation-based semantics, in: Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning, 2016, pp. 12–21.

[19] P. Matt, F. Toni, A game-theoretic measure of argument strength for abstract argumentation., in: Joint European Conference on Logics in Artificial Intelligence, 2008, pp. 285–297.

[20] E. Bonzon, J. Delobelle, S. Konieczny, N. Maudet, A comparative study of ranking-based semantics for abstract argumentation, in: Proceedings of AAAI'16, 2016, pp. 914–920.

[21] D. M. Gabbay, O. Rodrigues, An equational approach to the merging of argumentation networks, Journal of Logic and Computation 24 (6) (2014) 1253–1277.

[22] D. M. Gabbay, O. Rodrigues, Equilibrium states in numerical argumentation networks, Logica Universalis 9 (4) (2015) 411–473.

[23] J.Leite, J. Martins, Social abstract argumentation., in: Int. Joint Conference on Artificial Intelligence, 2011, pp. 2287–2292.

[24] G. Brewka, Preferred subtheories: An extended logical framework for default reasoning, in: Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89), 1989, pp. 1043–1048.

[25] D. N. Walton, Argument Schemes for Presumptive Reasoning, Lawrence Erlbaum Associates, 1996.

[26] H. Prakken, A study of accrual of arguments, with applications to evidential reasoning, in: Proceedings of the tenth International Conference on Artificial Intelligence and Law, ACM Press, 2005, pp. 85–94.

[27] J. L. Pollock, Defeasible reasoning, Cognitive Science 11 (1987) 481–518.

[28] V. Lifschitz, Foundations of logic programming, in: Principles of Knowledge Representation, CSLI Publications, 1996, pp. 69–127.

[29] Y. Venema, Lectures on the modal μ-calculus, renmin University in Beijing (China) (2008).

[30] R. Baumann, C. Spanring, Infinite argumentation frameworks, in: Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation, Vol. 9060 of LNCS, 2015, pp. 281–295.

[31] S. Modgil, (formal) argumentation theory, The Reasoner (`www.thereasoner.org`) 11 (6) (June, 2017) 32–34.

[32] D. van Dalen, Logic and Structure, Springer, 1980.

[33] H. Prakken, An abstract framework for argumentation with structured arguments, Argument and Computation 1 (2010) 93–124.

[34] G. Brewka, Adding priorities and specificity to default logic, in: European Workshop on Logics in Artificial Intelligence, 1994, p. 247?260.

[35] A. P. Young, S. Modgil, O. Rodrigues, Prioritised default logic as rational argumentation, in: Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems, AAMAS '16, 2016, pp. 626–634.

[36] D. Walton, G. Sartor, Teleological justification of argumentation schemes, Argumentation 27 (2) (2013) 111–142.

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

[38] B. Verheij, Dialectical argumentation with argumentation schemes: An approach to legal logic, Artificial Intelligence and Law 11 (2-3) (2003) 167–195.

[39] I. Rahwan, M. I. Madakkatel, J. Bonnefon, R. N. Awan, S. Abdallah, Behavioral experiments for assessing the abstract argumentation semantics of reinstatement, Cognitive Science (8) (2010) 1483–1502.

[40] S. Modgil, F. Toni, F. Bex, I. Bratko, C. I. Chesñevar, W. Dvořák, M. A. Falappa, X. Fan, S. Gaggl, A. J. García, M. P. González, T. F. Gordon, J. Leite, M. Možina, C. Reed, G. R. Simari, S. Szeider, P. Torroni, S. Woltran, Chapter 21: The added value of argumentation, in: S. Ossowski (Ed.), Agreement Technologies, Springer, 2013, pp. 357–403.

[41] K. Atkinson, What Should We Do?: Computational Representation of Persuasive Argument in Practical Reasoning, PhD Thesis, Department of Computer Science, University of Liverpool, Liverpool, UK, 2005.

[42] L. Amgoud, S. Vesic, Handling inconsistency with preference-based argumentation, in: Scalable Uncertainty Management: 4th International Conference, SUM 2010, 2010, pp. 56–69.

[43] P. M. Thang, H. T. Luong, Translating preferred subtheories into structured argumentation, Journal of Logic and Computation 24 (4) (2014) 831–849.

[44] S. Modgil, Technical report: Argumentative characterisations of non-monotonic inference in preferred subtheories: Stable equals preferred, `https://nms.kcl.ac.uk/sanjay.modgil/StableEqualsPreferred.pdf`.

[45] D. Gabbay, O. Rodrigues, Degrees of "in", "out" and "undecided" in argumentation networks., in: Proc. Computational Models of Argument, 2016, pp. 319–326.

[46] M. Caminada, L. Amgoud, On the evaluation of argumentation formalisms, Artificial Intelligence 171 (56) (2007) 286 – 310. `doi:http://dx.doi.org/10.1016/j.artint.2007.02.003`.
URL `http://www.sciencedirect.com/science/article/pii/S0004370207000410`

[47] S. Modgil, M. Caminada, Proof theories and algorithms for abstract argumentation frameworks, in: I. Rahwan, G. Simari (Eds.), Argumentation in Artificial Intelligence, Springer, 2009, pp. 105–132.

[48] D. Grossi, Abstract argument games via modal logic, Synthese 190 (1) (2013) 5.