

On the Graded Acceptability of Arguments

Davide Grossi

Department of Computer Science
University of Liverpool
d.grossi@liverpool.ac.uk

Sanjay Modgil

Department of Informatics
King's College London
sanjay.modgil@kcl.ac.uk

Abstract

The paper develops a formal theory of the degree of justification of arguments, which relies solely on the structure of an argumentation framework. The theory is based on a generalisation of Dung's notion of acceptability, making it sensitive to the numbers of attacks and counter-attacks on arguments. Graded generalisations of argumentation semantics are then obtained and studied. The theory is applied by showing how it can arbitrate between competing preferred extensions and how it captures a specific form of accrual in instantiated argumentation.

1 Introduction

A Dung argumentation framework [Dung, 1995] relates arguments by binary conflict-based attacks. The sceptically justified arguments are those in the intersection (respectively union, if credulously justified) of sets (extensions) of 'acceptable' arguments evaluated under various semantics (see [Baroni *et al.*, 2011] for an overview). The arguments and attacks may be assumed to be defined by an instantiating set of logical formulae (theory); the claims of justified arguments thus identifying the instantiating theory's non-monotonic inferences. Extensions are evaluated based on the principle that if an argument a is not attacked, then a is justified. If a is attacked by an un-attacked b , then b is justified, a rejected. If an un-attacked c attacks b , then b is rejected, and c reinstates a as justified (c defends a), and so on.

Context and objectives In prominent developments of [Dung, 1995], an attack from b to a may not succeed, if a is preferred to b (e.g., [Amgoud and Cayrol, 2002; Modgil and Prakken, 2013]), or because an attack's weight may fall below a numerical threshold [Dunne *et al.*, 2011]. In the so-called equational approaches (e.g., [Egilmez *et al.*, 2013; Gabbay and Rodrigues, 2012]) attacks succeed to a given quantifiable degree, based on the propagation of numerical values assigned to arguments and/or attacks, which then determine the degree to which arguments are justified. The above developments rely on exogenously given qualitative or quantitative information to decide, or give more refined accounts of, the justification status of arguments. However, one can intuitively refine the notion of justification based on *the number*

of attacks on an argument. For example, consider an argument g , which consists of two sub-arguments concluding that a suspect had *opportunity* and *motive*, and defeasibly extends these sub-arguments to conclude the presumed guilt of a suspect. The level to which g (and the presumption of guilt) can be said to be justified, is intuitively lessened in the case that counter-arguments to the suspect's *motive* and *opportunity* can be established (i.e., g is attacked on the conclusions of both its sub-arguments) as compared with a counter-argument to only one of *motive* or *opportunity* (g is attacked on only one sub-argument).

This paper sets out to generalise Dung's theory so as to differentiate levels of justification of arguments, by relying only on the structure of the Dung graph, and *without* relying on any exogenous information. We do so by formalising the two following principles. Everything else being equal, (A1) the greater the number of attacks on an argument b , the weaker is b 's level of justification, in which case the attacking arguments can be said to 'accrue' in weakening b . Thus if b is attacked by c and c' , both un-attacked, then b has a lower justification than when attacked by c alone. This in turn means that if a is attacked by b , then a is more strongly justified when defended by c and c' 's attacks on the weakened b , than when a is defended by c alone. Hence, all else being equal, (A2) the larger the number of arguments defending a , the stronger is a 's level of justification, in which case the defending arguments can be thought of as 'accruing' in strengthening a .

Outline of the paper In Section 2 we review Dung's theory. Section 3 then generalises Dung's notion of the defense/acceptability of a by a set of arguments X , by grading defense with respect to the number of arguments attacking and defending a . This yields a ranking among types of 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 with respect to the existence and computation problems. This makes it possible to rank arguments according to how strongly they are justified under different graded semantics. Section 5 then shows two applications of the graded semantics to arbitrate between multiple preferred extensions, and to ASPIC+ argument instantiations. Section 6 concludes and compares our work with other formalisms ranking arguments

based only on their interactions [Besnard and Hunter, 2001; Cayrol and Lagasque-Schieux, 2005; Matt and Toni, 2008; Wu and Caminada, 2010; Amgoud and Ben-Naim, 2013]. The section argues that our approach is the first to generalise all the standard concepts and semantics of Dung’s theory to novel graded variants, yielding a more fine grained notion of acceptability-based justification. Proofs do not pose particular challenges and make use of standard techniques (e.g., fixpoint theory). They are omitted for space reasons.

2 Abstract Argumentation

We review Dung’s abstract argumentation theory [Dung, 1995] in terms of its fixpoint theory presentation, as this provides a natural set up for pursuing the objectives of the paper.

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 $a \rightarrow b$ denotes that a attacks b , and $X \rightarrow a$ denotes that $\exists b \in X$ s.t. $b \rightarrow a$. Similarly, $a \rightarrow X$ denotes that $\exists b \in X$ s.t. $a \rightarrow b$. An AF such that for each $a \in \mathcal{A}$ the cardinality $|\{b \mid a \leftarrow b\}|$ of the set of attackers of a is finite, is called finitary.*

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 function, applied to some $X \subseteq \mathcal{A}$, returns the arguments defended by X [Dung, 1995]. We call this function the defense function.

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

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

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 in [Pollock, 1987], and we refer to it as neutrality function.

Definition 3 (Neutrality Function). *The neutrality function $n_\Delta : \wp(\mathcal{A}) \rightarrow \wp(\mathcal{A})$ for $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ is defined as follows. For any $X \subseteq \mathcal{A}$: $n_\Delta(X) = \{x \in \mathcal{A} \mid \text{NOT } X \rightarrow x\}$.*

One can define the extensions, and hence the justified arguments, of an AF Δ under Dung’s semantics, in terms of the fixpoints ($X = d_\Delta(X)$ and $X = n_\Delta(X)$) or post-fixpoints ($X \subseteq d_\Delta(X)$ and $X \subseteq n_\Delta(X)$) of the defense and neutrality functions as follows:¹

Definition 4 (Semantics). *Let $\Delta = (\mathcal{A}, \rightarrow)$, $X \subseteq \mathcal{A}$. Then: X is a conflict-free extension of Δ iff $X \subseteq n_\Delta(X)$; X is an admissible extension of Δ iff X is conflict-free and $X \subseteq d(X)$; X is a complete extension of Δ iff $X \subseteq n(X)$ and $X = d(X)$; X is a stable extension of Δ iff $X = n(X)$; X is the grounded extension of Δ iff X is the smallest complete extension of Δ ; X is a preferred extension of Δ iff X is a largest complete extension of Δ .*

For semantics $S \in \{\text{complete, grounded, stable, preferred}\}$, $a \in \mathcal{A}$ is credulously, respectively sceptically, justified under S , if a is in at least one, respectively all, S extensions of Δ .

¹Henceforth, the subscript Δ indexing d and n will be omitted.

Finally, we recapitulate some well-known properties of the defense and neutrality functions that will be of use later:

Fact 1. [Dung, 1995] *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF and $X, Y \subseteq \mathcal{A}$. The following holds:*

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

Recall also that for a finitary AF, iterating (up to ω) the application of d_Δ from \emptyset yields the least fixed point of d_Δ (grounded extension) [Dung, 1995].

3 Graded Defense

In argumentation the justified status of an argument depends on the status of its attackers, which in turn depends on the status of their attackers, and so on. The recursive nature of argumentation semantics is neatly captured by the fixpoint constructs of Definition 4. We now refine this notion of status building on the simple assumptions that, all else being equal: **(A1)** having fewer attackers is ‘better’ than having more; **(A2)**, having more defenders is ‘better’ than having fewer. It is important to observe right away that Dung’s semantics are consistent with these assumptions as having no attackers is a sufficient condition for being justified under any semantics, and having no defenders is a sufficient conditions for being rejected under any semantics.

We proceed as follows. First we define a generalisation of the notion of defense that accounts for the above assumptions. Then we define variants of Dung’s semantics and of the notion of justifiability based on this new notion of defense. Finally we use these semantics to ‘rank’ arguments in a way that does not use exogenous information (such as preferences or weights) and only relies on the argumentation framework.

3.1 Graded Defense and Neutrality Functions

Let us illustrate our guiding intuitions with a few examples. Keep in mind that an argumentation framework represents all information available for the evaluation of the status of its arguments.

Example 1. *In Figures 1i) – 1iv), the encircled set X_i defends a_i ($i = 1 \dots 4$) under Dung’s Definition 2. However we can differentiate all these cases based on the number of attackers and defenders of a_i . For instance, X_2 more strongly defends a_2 in ii) than X_1 defends a_1 in i), as a_2 is defended by two arguments whereas a_1 is defended by one argument. Hence by **A2**, a_2 has a ‘stronger status’ than a_1 . However, illustrating **A1**, a_3 in iii) is attacked by two arguments as compared with the single attack on a_1 . Hence a_3 has lower status than a_1 . Note that each of a_1 and a_3 are defended against each individual attack by a single defending argument; hence the number of attackers differentiates between the two cases.*

Now, take an argument a and a set of arguments X . Let m be the number of a ’s attackers (b_1, \dots, b_m) and, for each b_i ($0 \leq i \leq m$) let n_i be the number of attackers of b_i (i.e., counter-attackers of a) in X . Finally, let n be the minimum among the n_i s, i.e., $n = \min(\{n_i\}_{0 \leq i \leq m})$. So all attackers of a have at least n counter-attackers, and integers m and n

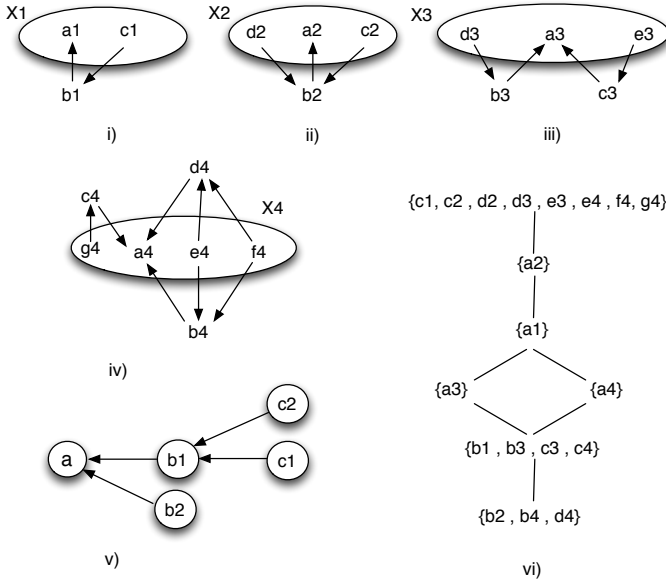


Figure 1: i)–iv) $X_{i \in \{1..4\}}$ defending a_1, \dots, a_4 ; v) example framework; vi) Hasse diagram ranking arguments in i)–iv).

encode information about how strongly a is defended by 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 integers. Then:*

$$d_n^m(X) = \{x \mid \exists_{\geq m} y : [y \rightarrow x \text{ AND } \exists_{\geq n} z : z \rightarrow y \text{ AND } z \in X]\}$$

where $\exists_{\geq n} x$, for integers n ('there exists at least n x ') are the standard first-order logic counting quantifiers.

So, $d_n^m(X)$ is the set of arguments which do not have at least m attackers (i.e., which have at most $m - 1$ attackers) that are not counter-attacked by at least n arguments in X . To illustrate, in Figure 1, $a_1 \in d_1^1(X_1)$ and $a_2 \in d_1^1(X_2)$ since in both cases: it is not the case that one or more arguments attacking a_1 , respectively a_2 , are not attacked by one or more defending arguments in X_1 (X_2). However if we increment n by 1: $a_2 \in d_2^1(X_2)$ but $a_1 \notin d_2^1(X_1)$ since it is the case that one or more arguments attacking a_1 are not attacked by two or more arguments in X_1 .

Definition 6 (Graded neutrality function). *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF and let m be any integer. Then: $n_m(X) = \{x \in \mathcal{A} \mid \exists_{\geq m} y : y \rightarrow x \text{ AND } y \in X\}$.*

The following fact shows that graded defense and neutrality are well-behaved generalisations of the standard defense and neutrality (Section 2) as they inherit their properties.

Fact 2. *For any AF $\Delta = \langle \mathcal{A}, \rightarrow \rangle$, the following holds:*

$$\begin{aligned} d_1^1(X) &= d(X) & (1) \\ n_1(X) &= n(X) & (2) \\ X \subseteq Y &\implies n_m(Y) \subseteq n_m(X) & (3) \\ X \subseteq Y &\implies d_n^m(X) \subseteq d_n^m(Y) & (4) \\ n_m(n_n(X)) &= d_n^m(X) & (5) \end{aligned}$$

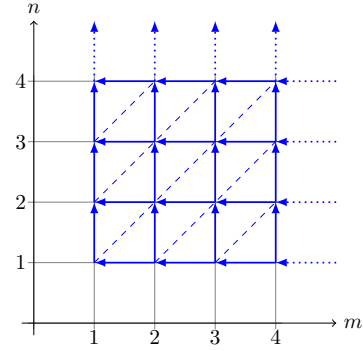


Figure 2: Partial ordering on graded defense functions (with $0 < m, n$). The horizontal and vertical axes consist of the values of m and, respectively, n . Arrows go from worse to better functions. Dashed lines (diagonals) denote incomparability.

Equation (1) reformulates $d(X)$ 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 (2) for $n(X)$. The remaining formulae generalise Fact 1 to the graded setting.

Fact 3. *For any AF $\Delta = \langle \mathcal{A}, \rightarrow \rangle$, $X \subseteq \mathcal{A}$, and $m, n < \omega$:*

$$n_m(X) \subseteq n_{m+1}(X) \quad (6)$$

$$d_n^m(X) \subseteq d_{m+1}^n(X) \quad (7)$$

$$d_n^m(X) \supseteq d_{n+1}^m(X) \quad (8)$$

Intuitively, Formula (6) states that the set of arguments attacked by at most m arguments in X is included in the set of arguments attacked by at most $m + 1$ arguments in X . Formulae (7) and (8) result from combining this with the fact that d_n^m is the composition of n_m with n_n (Formula (5)).

3.2 Ranking Graded Defense Functions

With the notion of graded defense we can give a mathematically precise formulation of our two driving assumptions. Fix a set of arguments X .

- **(A1):** The less attackers that are not counter-attacked by n defending arguments in X the better (for some fixed n). Therefore, for $m < t$ we have that belonging to $d_n^m(X)$ is 'better', w.r.t. X , than belonging to $d_n^t(X)$.
- **(A2):** For all m attackers (for a fixed m), the more counter-attacks from arguments in X , the better. Therefore, for $t < n$ we have that belonging to $d_n^m(X)$ is 'better', w.r.t. X , than belonging to $d_t^m(X)$.

Graded defense functions can be viewed as different 'types' or 'standards' of defense/acceptability. They can be ordered according to the following relation:

Definition 7. *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF. Define:*

$$d_n^m \triangleright d_t^s \iff m \leq s \text{ AND } t \leq n \quad (9)$$

The relation is depicted in its generality in Figure 2. Expressions $d_m \triangleright d_s$ may be read as follows: ‘being $\frac{m}{n}$ -defended is preferable over being $\frac{s}{t}$ -defended’ or ‘the $\frac{m}{n}$ -defense function is at least as strong as the $\frac{s}{t}$ -defense function.’

Fact 4. Relation \triangleright is: i) reflexive, antisymmetric and transitive (a partial order); ii) and if $d_m \triangleright d_s$, then for all $X \subseteq \mathcal{A}$, $d_m(X) \subseteq d_s(X)$.

Claim ii) states that being defended by a stronger defense function is logically more demanding than being defended by a weaker one. This is in line with the intuition that if an argument meets a demanding standard of defense it also meets a less demanding one.

Referring to the framework in Fig.1v), we illustrate Formula (7) and Fact 4: $d_1(\{c1, c2\}) = \{c1, c2, b2\} \subseteq d_2(\{c1, c2\}) = \{c1, c2, a, b2\} \subseteq d_3(\{c1, c2\}) = \{c1, c2, a, b1, b2\}$. We also illustrate Formula (8) and Fact 4 with reference to Figure 1ii): $d_1(\{d2, c2, a2\}) = \{d2, c2, a2\} \supseteq d_2(\{d2, c2, a2\}) = \{d2, c2, a2\} \supseteq d_3(\{d2, c2, a2\}) = \{d2, c2\}$.

Since the relation \triangleright is a partial order some functions may be incomparable (see Figure 2) and this, we claim, is intuitive.

Example 2. (Example 1 continued) In iv), $a4$ has three attackers, and so by **A1** $a3$ in iii) would be stronger than $a4$. On the other hand, every attacker of $a3$ is defended by only one argument, whereas two attacks on $a4$ are defended by two arguments, and so by **A2**, $a4$ would be stronger than $a3$. So $a3$ and $a4$ appear to be incomparable.

Further principles along **A1** and **A2** could be assumed in order to force \triangleright to become total. A natural one, which we refrain from studying in detail in this paper, is (**A3**): if a set of arguments is $\frac{m}{n}$ -defended and another one is $\frac{s}{t}$ -defended, where $m < s$ and $n < t$ (i.e., they are incomparable w.r.t. \triangleright) then the first one is more strongly defended because it has fewer attackers. Therefore, for $m < s$ and $n < t$, belonging to $d_m(X)$ is ‘better’ than belonging to $d_s(X)$. One could then redefine \triangleright as follows: $d_m \triangleright d_s$ iff either $m < s$ or, $m = s$ and $n \geq t$. This is nothing but a lexicographic order on graded defense functions, first based on their number of attackers, and then on their number of defenders.

3.3 Ranking Arguments by Graded Defense

The same recursive principles underpinning standard Dung’s 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 that $\frac{m}{n}$ -defends all its contained arguments, one can iteratively apply d_m to X .

Let us fix some terminology. For α ordinal, the α -fold iteration of d_m with m and n integers is denoted by d_m^α . For a given X , an infinite iteration generates an infinite sequence, or stream, $d_m^0(X), d_m^1(X), d_m^2(X), \dots$. A stream ‘stabilizes’ if and only if there exists an ordinal α such that $d_m^\alpha(X) = d_m^{\alpha+1}(X)$. Such a set $d_m^\alpha(X)$ is then called the limit of the stream. Since d_m is monotonic (Fact 2) such a

limit always exists by the Knaster-Tarski theorem. In finitary argumentation frameworks a limit is reachable in a countable number of steps and we can define the indefinite iteration of function d_m as: $d_m^*(X) = \bigcup_{0 \leq i < \omega} d_m^i(X)$.

Iterated graded defense can thus rank arguments:

Definition 8. 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:

$$a \succeq_X b \iff \forall m, n \geq 0 \text{ IF } b \in d_m^*(X) \text{ THEN } a \in d_n^*(X)$$

The strict part \succ_X is defined in the usual way.

The key intuition behind this definition is the following one. 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 4). By Definition 8 it is easy to see that \succeq_X is a partial order.

Example 3. Let us rank, by iterated defense w.r.t \emptyset , the arguments in Figure 1i-iv). Applying Definition 8 we obtain the partial order shown in Figure 1vi). Note that we assume one single AF Δ consisting only of the arguments and attacks shown in Figures 1i-iv). Under standard Dung’s 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). We can differentiate amongst these arguments. As expected the best arguments are those with no attackers. Second-best is $a2$ whose attackers are all counter-attacked by two un-attacked arguments. Then $a1$, since $a2 \in d_2^*(\emptyset)$ but $a1 \notin d_1^*(\emptyset)$ (and so $a1 \not\succeq_\emptyset a2$), but $\forall m, n$: if $a1 \in d_m^*(\emptyset)$ then $a2 \in d_n^*(\emptyset)$ (and so $a2 \succeq_\emptyset a1$). We then have that $a3$ and $a4$ are incomparable (cf. Example 2). Formally, $a3 \in d_3^*(\emptyset)$ and $a3 \notin d_2^*(\emptyset)$, but $a4 \notin d_3^*(\emptyset)$ and $a4 \in d_2^*(\emptyset)$. Critically, we can also differentiate amongst the rejected arguments (those not in the Dung grounded extension). Thus $b1, b3, c3, c4$ are ranked above $b2, b4, d4$.

Definition 8 allows one to rank arguments by the quality of the iterated defense they obtain from a given set. We will see that, depending on the nature of the given set X , and the relative size of n and m , Definition 8 relates in a precise way to the ranking of arguments w.r.t. graded generalisations of Dung’s semantics, to which we turn in the next section.

4 Graded Semantics

4.1 Generalising Dung’s semantics

We now generalise Definition 4 as follows:

Definition 9 (Graded Extensions). Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF, $X \subseteq \mathcal{A}$, and m and n be integers. Then: X is m -conflict-free iff $X \subseteq \mathfrak{n}_m(X)$; X is an mn -admissible extension of Δ iff $X \subseteq \mathfrak{n}_m(X)$ and $X \subseteq d_n^*(X)$; X is an mn -complete extension of Δ iff $X \subseteq \mathfrak{n}_m(X)$ and $X = d_n^*(X)$; X is an m -stable extension of Δ iff $X = \mathfrak{n}_m(X)$; X is the mn -grounded extension of Δ iff X is the smallest mn -complete extension of Δ ; X is an mn -preferred extension of

Δ iff X is a largest mn -complete extension of Δ .
For $S \in \{mn\text{-complete}, mn\text{-grounded}, m\text{-stable}, mn\text{-preferred}\}$, $a \in \mathcal{A}$ is credulously, respectively sceptically justified under semantics S , if a is in at least one, respectively all, S extensions of Δ .

We henceforth assume the sceptical definition when referring to an argument as being justified. For a given semantics, we can rank the justification status of an argument with respect to a given framework, exactly like we did in Definition 8 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 $S \in \{mn\text{-complete}, mn\text{-grounded}, m\text{-stable}, mn\text{-preferred}\}$:*

$$a \succeq_{\Delta}^S b \iff \forall m, n \geq 0, \text{ IF } b \text{ is justified w.r.t. } S \\ \text{ THEN } a \text{ is justified w.r.t. } S.$$

We showcase applications of this definition in Section 5.

4.2 Existence and computation

We focus now on the two following questions: do the semantics of Definition 9 always exist? And how can they be computed? Firstly, we state that a variant of Dung’s fundamental lemma [Dung, 1995] holds for graded admissibility.

Theorem 1 (Fundamental lemma). *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be an AF, and assume: (1) $X \subseteq \mathcal{A}$ is an mn -admissible extension of Δ ; (2) $x \in \mathbf{d}_m^n(X)$; (3) $n \geq m$. Then $X \cup \{x\}$ is mn -admissible.*

The assumption $n \geq m$ plays a crucial role in the proof of Theorem 1, which we omit for space reasons, and indicates that some graded semantics might not always exist if $n < m$. By means of example, assume the AF Δ consisting only of the arguments and attacks in Figure 3(iii). Then Δ has no 21-complete (and hence no 21-grounded or 21-preferred) extensions. To see this, observe that $\mathbf{d}_2^1(\emptyset) = X = \{a3, b3, c3, d3, e3\}$, but $X \not\subseteq \mathbf{n}_2(X)$ since X contains two attacks on $a3$, that is, X is not 2-conflict-free.

However, provided that $n \geq m$ the standard arguments for the existence of complete, grounded and preferred extensions [Dung, 1995] carry over to their graded variants (based on Fact 1). In particular, Fact 1 implies that if $n \geq m$, then the mn -admissible extensions form a complete partial order² w.r.t. set inclusion, and so every admissible extension is a subset of a maximal under set inclusion mn -admissible (i.e., mn -preferred) extension (by Zorn’s lemma, cf. [Baumann and Spanring, 2015]). Hence if X is mn -admissible, then checking whether $x \in \mathbf{d}_m^n(X)$ amounts to checking whether x belongs to a mn -preferred extension. In addition, Fact 1 gives us some ground to sketch how graded semantics can be computed by fixpoint approximation:

Theorem 2. *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be a finitary AF, $X \subseteq \mathcal{A}$ be an mn -admissible extension and $n \geq m$. Then $\mathbf{d}_m^n(X)$ is the smallest mn -complete extension containing X .*

Corollary 1. *Let $\Delta = \langle \mathcal{A}, \rightarrow \rangle$ be a finitary AF, and $n \geq m$. Then $\mathbf{d}_m^n(\emptyset)$ is the mn -grounded extension of Δ .*

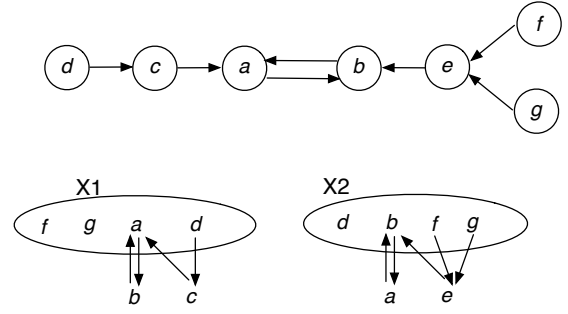


Figure 3: An AF with two preferred extensions

Corollary 1 allows us to compute the grounded extension when $n \geq m$. Also, Theorem 2 implies that if we choose a ‘large enough’ mn -admissible set X then $\mathbf{d}_m^n(X)$ will yield an mn -preferred extension when $n \geq m$. Referring to Figure 3’s AF: $\mathbf{d}_1(\emptyset) = \{d, f, g\}$; $\mathbf{d}_2(\{d, f, g\}) = \{d, f, g\}$. Hence $\{d, f, g\}$ is the 12-grounded extension.

The above results suggest that among Definition 9’s graded semantics, only those for which $n \geq m$ are *well-behaved* in that they retain the basic characteristics—dependent on the fundamental lemma—of Dung’s original semantics. This is an interesting finding and fits well with our guiding intuition that one obtains higher justifiability when defenders are many and attackers few. When this ratio is inverted, with attackers outnumbering defenders, then justifiability criteria (extensions) may altogether cease to exist.

Note finally that Theorem 2 reveals a precise relation between the ranking of arguments w.r.t. iterated application of \mathbf{d}_m^n to an admissible X in Definition 8, and the ranking of arguments w.r.t. well-behaved graded variants of Dung’s semantics in Definition 10, given that iterated applications of graded defense yield the graded extensions.

5 Applying Graded Semantics

We now illustrate applications of the framework for deciding between preferred extensions of abstract frameworks, and for instantiations of the structured ASPIC+ framework.

Firstly, consider Figure 3. Its AF has two preferred extensions— $X1$ and $X2$ —under the standard Dung semantics. This equates with $X1$ and $X2$ both being 11-preferred extensions in the graded terminology. However, only $X2$ is subsumed by a 12-preferred extension.³ Hence b is justified under 12-preferred whereas a is not. Indeed, it is the case that $\forall m, n \geq 0$, if a is justified under mn -preferred, so is b , but not vice versa. Hence $b \succ_{\Delta}^{\text{preferred}} a$. We can thus use graded semantics to arbitrate between arguments. Traditionally, an exogenously given preference [Amgoud and Cayrol, 2002; Modgil and Prakken, 2013] is used to arbitrate between such arguments, whereas we show one can employ information that is already available in standard AFs for that purpose.

We now consider instantiations of the ASPIC+ [Modgil and Prakken, 2013] framework. ASPIC+ formalises logic-

²See [Davey and Priestley, 1990, Ch. 8] for a definition.

³That is, $X2$ is 12-admissible whereas $X1$ is not.

based accounts of reasoning with schemes and critical questions [Walton, 1996]. For example, consider an argument c for ‘invading Syria’ which reasons that since ‘Assad is suppressing Syrians’, invasion will ‘remove Assad from power’, achieving ‘Syrian democracy’ and so promoting ‘peace’. Consider then the critical questions addressed by two ASPIC+ arguments $a1$ and $a2$ (in their simplified propositional representation below) attacking argument c .

Specifically, let $\{e1, e2, e3, s1, s2, s3\}$ be premises and $e1, s1 \Rightarrow \neg g$; $e2, s2 \Rightarrow \neg g$; $e3, s3 \Rightarrow g$ defeasible inference rules, where $e1, e2$ and $e3$ respectively denote that the Institute for War Studies (WS), Arab League (AL), and United Nations (UN) are experts on middle east affairs. $s1$ ($s2$) denotes that ‘the WS (AL) state that removing Assad will not achieve democracy’, $s3$ denotes that ‘the UN states that removing Assad will achieve democracy’, and g denotes that ‘removing Assad will achieve democracy’. We then have the $AF \Delta$ with the following arguments (we ignore representations of c and the set of arguments X built from the premises alone, such as $[e1], [s1], [e2]$ etc):

$a1 = [e1; s1; e1, s1 \Rightarrow \neg g]$ and $a2 = [e2; s2; e2, s2 \Rightarrow \neg g]$ and $a3 = [e3; s3; e3, s3 \Rightarrow g]$, and attacks: $a1 \leftrightarrow a3 \leftrightarrow a2$.

Now recall the introductory ‘guilty suspect’ example of Section 1. In that example two attacks on *distinct* premises accrue to weaken the argument for guilt. In the current example, $a1$ and $a2$ attack c on the same premise (assumption⁴), and (as illustrated with the rejected arguments in Example 3) is weaker when attacked by both $a1$ and $a2$, than by only one of $a1$ or $a2$. c ’s weakened status can be seen to account for the *accrual of arguments* (i.e., mutual strengthening) of $a1$ and $a2$ in support of their common claim. Indeed, Δ has two preferred extensions: $E1 = \{a1, a2\} \cup X$ and $E2 = \{a3, c\} \cup X$. However, under the graded semantics we obtain (abusing notation) $a1, a2 \succ_{\Delta}^{preferred} a3, c$. Intuitively, $a1$ and $a2$ are each defended by two arguments, whereas $a3$ and c are each defended by only one argument.

The incompatibility of accrual with Dung’s standard theory has been discussed and argued for in [Prakken, 2005]. Our graded notion of acceptability partially challenges this view, showing how a simple counting-based form of accrual can be coherently accommodated within Dung’s theory.

6 Conclusions

Related work A few other works rank arguments based solely on attacks. [Besnard and Hunter, 2001] also formalise intuitions similar to **A1** and **A2**. However they do not investigate these ideas in the context of Dung’s theory, but do so only for acyclic graphs with deductive arguments. Evaluation of arguments is achieved quantitatively through a function (called a *categoriser*) $v : \mathcal{A} \rightarrow [0, 1]$ defined as $v(a) = \frac{1}{1 + \sum_{a \rightarrow b} v(b)}$, 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 Lagasque-Schiex, 2005] then

⁴ASPIC+ formalises asymmetric attacks on assumptions [Bodarenko *et al.*, 1997] (in this case the assumption premise g used in argument c).

generalises use of this function to Dung frameworks, and develop (in their terms) a ‘local approach’ to valuation of arguments that uses this function. A formal comparison between categorisers and rankings obtainable through Definition 10 (or variants thereof) is certainly worth investigating. However [Cayrol and Lagasque-Schiex, 2005] then formalise a ‘global approach’ that they argue gives more intuitive outcomes when ranking arguments. This approach requires a highly involved processing of cyclical graphs that does not lend itself to a systematic generalisation of Dung’s theory. Furthermore, the global approach yields outcomes incompatible with the intuition that the more the number of attacks the weaker the argument (i.e., **A1**). For example, given $AF1 = c1 \rightarrow b1 \rightarrow b, c2 \rightarrow b$ and $AF2 = c3 \rightarrow b'$, then [Cayrol and Lagasque-Schiex, 2005] argue that b should be ranked higher than b' since “it has at least a defense whereas b' has none”. However our approach ranks b' higher than b (both are d_1^1 justified, but only b' is d_n^2 justified for $n > 1$). Intuitively, both b and b' are attacked by a single un-attacked argument ($c2$ and $c3$ respectively). But then b is also attacked by $b1$, and even though $c1$ defends this attack, b is (by **A1**) weaker than when un-attacked (as illustrated also by the comparative ranking of $a1$ and the un-attacked arguments in Figure 1vi)). Indeed, more recently, [Amgoud and Ben-Naim, 2013] also formalise the intuition articulated by **A1**, as well as the intuition that an argument retains a degree of justification when attacked by an un-attacked argument. They develop two ranking-based semantics, which they call discussion-based and burden-based, satisfying some rationality principles they argue for. These semantics are also based on the processing of attack paths and their applicability relies on conjectures concerning the processing of cyclical paths. How they relate to the standard Dung’s semantics is unclear.

Finally, [Matt and Toni, 2008] and [Wu and Caminada, 2010] also present accounts of the degree to which an argument is acceptable. The former develops a very different approach to ours, as it defines argument strength in terms of the value of a repeated two-person zero-sum strategic game with imperfect information. The latter refines the standard notion of justifiability by isolating six different statuses of arguments based on combinations of standard notions from Dung’s theory (such as, 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) rather than incorporating an idea of graduality in those notions.

To conclude we stress that the key distinguishing feature of our work with respect to the above ones is that we ‘grade’ the justification status of arguments through a generalisation of the body of notions and techniques of Dung’s theory, such as defense/acceptability and extensions under all the standard semantics. None of the above mentioned approaches develops a comparable generalisation.

Conclusions and future work We have parameterised the defense of an argument by the number of its attackers and defenders. This yields rankings of defense functions and arguments generalising the standard acceptability-based semantics. We have applied these graded semantics to an instantiated AF and shown how this integrates a simple form of accrual with Dung’s theory. Other features of accrual [Prakken,

2005] are not captured by graded semantics, and a natural question concerns the extent to which graded semantics can shed light on the phenomenon of accrual. We will also look to the development of graded defense functions and semantics that capture further or alternative assumptions to **A1** and **A2**. For example, consider the principle **A3** proposed in Section 3.2. Also, some 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. We believe the technical notions introduced in this paper can be readily adapted to capture these alternative intuitions under the same framework. Future work should also evaluate our graded semantics with respect to rationality postulates for instantiated Dung frameworks [Caminada and Amgoud, 2007], and postulates for ranking semantics proposed in [Amgoud and Ben-Naim, 2013].

We conclude with a suggestive remark. The success of Dung’s theory lies in its foundational approach to both computational *and* (together with various extensions of the theory) human argumentation [Modgil *et al.*, 2013]. Regarding the latter, [Rahwan *et al.*, 2010] report that human subjects have higher confidence in the claims of arguments that are un-attacked, than when these arguments are subsequently attacked and then defended. This finding fits with our graded semantics assigning higher status to un-attacked arguments than to defended arguments.

Acknowledgments

We wish to thank the reviewers for their helpful comments. They greatly improved the paper.

References

- [Amgoud and Ben-Naim, 2013] L. Amgoud and J. Ben-Naim. Ranking-based semantics for argumentation frameworks. In *Scalable Uncertainty Management*, pages 134–147. Springer, 2013.
- [Amgoud and Cayrol, 2002] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.
- [Baroni *et al.*, 2011] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(4):365–410, 2011.
- [Baumann and Spanring, 2015] R. Baumann and C. Spanring. Infinite argumentation frameworks. In *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation*, volume 9060 of *LNCS*, pages 281–295, 2015.
- [Besnard and Hunter, 2001] P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128(1–2):203–235, 2001.
- [Bodarenko *et al.*, 1997] A. Bodarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.
- [Caminada and Amgoud, 2007] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(56):286 – 310, 2007.
- [Cayrol and Lagasquie-Schiex, 2005] C. Cayrol and M.-Ch. Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.
- [Davey and Priestley, 1990] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 1990.
- [Dung, 1995] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.
- [Dunne *et al.*, 2011] P. Dunne, T. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Weighed argument systems. basic definitions, algorithms and complexity results. *Artificial Intelligence*, 175(2):457–86, 2011.
- [Egilmez *et al.*, 2013] S. Egilmez, J. Martins, and J. Leite. Extending social abstract argumentation with votes on attacks. In *Proc. 2nd Int. Workshop on Theory and Applications of Formal Argumentation*, pages 16–31, 2013.
- [Gabbay and Rodrigues, 2012] D. M. Gabbay and O. Rodrigues. A numerical approach to the merging of argumentation networks. In *Proc. CLIMA XIII*, pages 195–212, 2012.
- [Matt and Toni, 2008] P. Matt and F. Toni. A game-theoretic measure of argument strength for abstract argumentation. In *Joint European Conference on Logics in Artificial Intelligence*, pages 285–297, 2008.
- [Modgil and Prakken, 2013] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195(0):361 – 397, 2013.
- [Modgil *et al.*, 2013] S. Modgil, F. Toni, and *et al.* Chapter 21: The added value of argumentation. In S. Ossowski, editor, *Agreement Technologies*, pages 357–403. Springer, 2013.
- [Pollock, 1987] J. L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.
- [Prakken, 2005] 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*, pages 85–94. ACM Press, 2005.
- [Rahwan *et al.*, 2010] I. Rahwan, M. I. Madakkatel, J. Bonnefon, R. N. Awan, and S. Abdallah. Behavioral experiments for assessing the abstract argumentation semantics of reinstatement. *Cognitive Science*, (8):1483–1502, 2010.
- [Walton, 1996] D. N. Walton. *Argument Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, 1996.
- [Wu and Caminada, 2010] Y. Wu and M. Caminada. A labelling-based justification status of arguments. *Studies in Logic*, 3(4):12–29, 2010.