Proceedings of the 22nd International Conference on Principles of Knowledge Representation and Reasoning
Main Track

On Gradual Semantics for Assumption-Based Argumentation

Anna Rapberger[1] , **Fabrizio Russo**[1] , **Antonio Rago**[1,2] and **Francesca Toni**[1]
[1]Department of Computing, Imperial College London, UK
[2]Department of Informatics, King's College London, UK
{a.rapberger, fabrizio, a.rago, ft}@imperial.ac.uk

Abstract

In computational argumentation, gradual semantics are fine-grained alternatives to extension-based and labelling-based semantics. They ascribe a dialectical strength to (components of) arguments sanctioning their degree of acceptability. Several gradual semantics have been studied for abstract, bipolar and quantitative bipolar argumentation frameworks (QBAFs), as well as, to a lesser extent, for some forms of structured argumentation. However, this has not been the case for assumption-based argumentation (ABA), despite it being a popular form of structured argumentation with several applications where gradual semantics could be useful. In this paper, we fill this gap and propose a family of novel gradual semantics for equipping assumptions, which are the core components in ABA frameworks, with dialectical strengths. To do so, we use bipolar set-based argumentation frameworks as an abstraction of (potentially non-flat) ABA frameworks and generalise state-of-the-art modular gradual semantics for QBAFs. We show that our *gradual ABA semantics* satisfy suitable adaptations of desirable properties of gradual QBAF semantics, such as *balance* and *monotonicity*. We also explore an argument-based approach that leverages established QBAF modular semantics directly, and use it as baseline. Finally, we conduct experiments with synthetic ABA frameworks to compare our gradual ABA semantics with its argument-based counterpart and assess convergence.

1 Introduction

Computational argumentation is a central research area within the field of knowledge representation and reasoning. Among the most prominent structured argumentation formalisms is *assumption-based argumentation (ABA)* (Bondarenko et al. 1997; Čyras et al. 2018), which provides a versatile framework for modelling and evaluating argumentative reasoning, with applications in several domains, such as recommendations within healthcare (Čyras et al. 2021; Skiba, Thimm, and Wallner 2024; Zeng et al. 2020) and explainable causal discovery (Russo, Rapberger, and Toni 2024). ABA frameworks (ABAFs) have been extensively studied, including investigations of their formal properties (Caminada and Schulz 2017; Berthold, Rapberger, and Ulbricht 2023), computational complexity (Čyras, Heinrich, and Toni 2021; Lehtonen, Wallner, and Järvisalo 2021a; Lehtonen et al. 2023) and implementations (Toni 2013; Lehtonen, Wallner, and Järvisalo 2021b; Lehtonen et al. 2024). The building blocks in ABA are *assumptions*, which are defeasible elements that can be challenged via a so-called *contrary function*, and *inference rules*. Relations between assumptions can be conflicting or supporting, where, informally: a set of assumptions S *attacks* an assumption a if the contrary of a is derivable (via a tree-derivation, see Section 3.1) from (a subset of) S; and S *supports* a if a is derivable from (a subset of) S.[1] To date, existing research in the evaluation of ABA frameworks focuses exclusively on two main strands. The first is *extension-based semantics* (Bondarenko et al. 1997), which use attack and support relations to decide which sets of assumptions can be jointly accepted. The second is *labelling-based semantics* (Schulz and Toni 2017), which categorise each assumption as accepted, rejected, or undefined. Both forms of semantics may be used to characterise multiple possible solutions for tasks such as decision making.

Gradual semantics is an orthogonal evaluation paradigm with growing popularity, most prominently in the realm of (quantitative, bipolar) abstract argumentation (Cayrol and Lagasquie-Schiex 2005; Baroni, Rago, and Toni 2019; Amgoud and Doder 2018). Abstract argumentation represents argumentative reasoning processes via directed graphs in which nodes correspond to abstract arguments and different edge types indicate, for example, attack (Dung 1995) and support (Amgoud et al. 2008) relations between arguments. In contrast to extension-based semantics, gradual semantics assign dialectical strengths to arguments, which can be seen as acceptability degrees, providing a finer-grained evaluation of arguments than the one given by extensions (Dung 1995; Cayrol and Lagasquie-Schiex 2005) or labellings (Caminada and Gabbay 2009; Gonzalez et al. 2021). In quantitative bipolar argumentation frameworks (QBAFs) (Baroni, Rago, and Toni 2019), gradual semantics compute a numerical *acceptability degree* for each argument based on its initial strength (the base score, which reflects the argument's intrinsic plausibility) and the acceptability degrees of its attackers and supporters. Gradual semantics have numerous applications in domains such as engineering decision support (Rago et al. 2016), social media analysis (Leite and Martins 2011; Oluokun et al. 2024), on-

[1]Supports emerge only if the ABAF is not *flat*, i.e., if assumptions are derivable (Bondarenko et al. 1997).

line review aggregation (Rago et al. 2025) and explainability (Ayoobi, Potyka, and Toni 2023), amongst others. Modular gradual semantics, originally proposed by Mossakowski and Neuhaus (2018), decompose these semantics into aggregation and influence functions, enabling a systematic study of their formal behaviour (see also (Baroni, Caminada, and Giacomin 2018; Potyka 2019; Potyka and Booth 2024)).

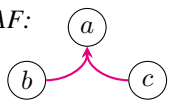
Despite the rich literature on both ABA and gradual semantics, there is currently no study investigating gradual semantics for ABA in contexts where assumptions themselves carry weights. This leaves a significant gap in the ABA theory, and limits its application exclusively to scenarios where multiple solutions are tolerable and a numerical evaluation of assumptions is not needed. However, many real-world scenarios require differentiating between assumptions based on weights that capture, for example, source credibility or empirical evidence (as, e.g., in the use of ABA for causal discovery (Russo, Rapberger, and Toni 2024)).

Example 1.1. *Consider a simple ABAF with inference rule $q \leftarrow b, c$, where a, b, c are assumptions, the contrary of a is q and the contraries of b and c are sentences x, y respectively. This ABAF may represent a very simple decision-making setting: The assumption a stands for 'having a picnic', q stands for 'having a picnic is a bad idea', b stands for 'bad weather is forecasted' and c stands for 'no friends are around'. Given that x, y are not derivable and thus b, c cannot be attacked, any extension- or labelling-based semantics will return the set $\{b, c\}$, indicating that a is not accepted. Suppose that, while only being able to treat b and c as assumptions (as they are about the future), we have some quantitative information (weights) about them, e.g., from a weather forecasting website and from the profile of the friends from a social network. If the weights of b and c are both high (because the website and profiles are both reliable), then we would expect a low degree of acceptability for a, but if they are both low, the degree should be higher.*

With the standard extension-based (Bondarenko et al. 1997) or labelling-based (Schulz and Toni 2017) ABA semantics, the numerical information available in such settings cannot be processed. In this paper, we take a foundational step towards closing this gap by exploring gradual semantics for *weighted assumptions* in ABA. To this end, we utilise *bipolar set-based argumentation frameworks* (BSAFs) (Berthold, Rapberger, and Ulbricht 2024), an abstract formalism that provides a concise representation for ABAFs by modelling the collective attacks and collective supports.

Example 1.1 (continued). *The BSAF representation contains a node for each assumption and a single attack $(\{b, c\}, a)$, depicted below (left). The table (right) contains possible values (base scores) for the assumptions.*

BSAF:



	a	b	c
τ_1	1	1	1
τ_2	1	0.1	0.2

Intuitively, the evaluation with respect to τ_1 should result in a much lower dialectical strength of assumption a than τ_2.

We employ BSAFs for their compactness and their flexibility to focus only on the defeasible elements of ABAFs, assumptions, whose granular evaluation is the focus of our proposed semantics. Our main contributions are as follows:

- We introduce a family of gradual semantics for ABA and investigate whether they satisfy fundamental properties of gradual semantics, such as balance and monotonicity, as well as their convergence behaviour. We introduce bespoke assumption aggregation functions and study their behaviour in relation to the other components of the modular semantics.
- Orthogonal to our (assumption-based) gradual ABA semantics, we craft an (argument-based) QBAF baseline method to evaluate an ABAF with weighted assumptions, building on established approaches. For this we leverage the correspondence between flat ABAFs and AFs to then employ existing gradual QBAF semantics.
- We conduct an experimental evaluation to study the convergence behaviour of our gradual ABA semantics and compare it to the QBAF baseline. Among our findings is that instantiations of the gradual ABA semantics converge in 90% of the scenarios while the QBAF baseline does so only in about 70% of the scenarios, demonstrating the advantages of our proposed native ABA approach on the existing AF-based pipeline. Code is provided at https://github.com/briziorusso/GradualABA.

All omitted proofs and experimental details can be found in the appendix of the technical report (Rapberger et al. 2025).

2 Related Work

The use of gradual semantics within structured argumentation frameworks in general has only recently started to receive attention. Within the context of ABA, one recent work (Skiba, Thimm, and Wallner 2023) explored the use of a family of ranking-based semantics for assumptions, reducing flat ABA frameworks to abstract argumentation and then applying existing ranking-based semantics, an endeavour which is orthogonal to our definition of purpose-built gradual semantics for general (possibly non-flat) ABA. To the best of our knowledge, there are no other works which explore gradual semantics within ABA.

Gradual semantics have also appeared in other forms of structured argumentation. Heyninck et al. (2023) explore the use of gradual semantics for *deductive argumentation* (Besnard and Hunter 2014). Here, the focus is on relationships between properties of gradual semantics and culpability (of any logical inconsistency) in abstract argumentation frameworks extracted from deductive argumentation. Further, in a restricted type of structured argumentation, namely *statement graphs* (Hecham, Bisquert, and Croitoru 2018), Jedwabny et al. (2020) define a modular gradual semantics based on complete support (inference) trees, demonstrating that one instance satisfies several useful properties centred around cases with complete information. Another study (Rago et al. 2024) explores the deployment of gradual semantics from abstract argumentation in these statement graphs, demonstrating that they present distinct advantages

in cases under incomplete information. Finally, Amgoud and Ben-Naim (2015) investigate ranking-based semantics for logic-based AFs, with no explicit quantitative strength.

Several works relate to ours in terms of aggregating the strengths of arguments. In the realm of ASPIC$^+$ (Modgil and Prakken 2014), Prakken (2024) provides a comprehensive study of argument strength, instantiated with a formal model, considering the different ways in which an argument can be attacked. Spaans (2021) explores initial argument strength aggregation functions in ASPIC$^+$, considering several principles similar in spirit to those in Section 4.1. Other works use probabilistic argument strength in the context of logic-based argumentation, as studied by e.g., Prakken (2018) in ASPIC$^+$ and Hunter (2022; 2013). Rossit et al. (2021) study accrual in quantitative abstract argumentation. In contrast to our work, they study accrual strengths in the context of extension-based semantics.

Lastly, ABA has been generalised to accommodate preferences (Čyras and Toni 2016) and priorities (Dung and Thang 2010; Fan 2023). These works incorporate numbers to evaluate acceptance. However, in contrast to gradual semantics, these approaches use preferences and priorities to modify the argumentation frameworks, e.g., by breaking symmetric attacks or by reversing attacks, to subsequently calculate extensions. Instead, gradual semantics propagate strengths across the entire framework with the final scores reflecting not only the initial (base) scores (which one could compare to preferences or priorities), but the whole structure.

3 Background

3.1 Assumption-Based Argumentation

We recall assumption-based argumentation (ABA) (Čyras et al. 2018). We assume a deductive system $(\mathcal{L}, \mathcal{R})$, where $\mathcal{L}$ is a formal language, i. e., a set of sentences, and $\mathcal{R}$ is a set of inference rules over $\mathcal{L}$. A rule $r \in \mathcal{R}$ has the form $a_0 \leftarrow a_1, \ldots, a_n$ with $a_i \in \mathcal{L}$. We write $head(r) = a_0$ and $body(r) = \{a_1, \ldots, a_n\}$ for the (possibly empty) body of r.

Definition 3.1. *An ABA framework (ABAF) is a tuple $(\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-)$ with deductive system $(\mathcal{L}, \mathcal{R})$, a non-empty set $\mathcal{A} \subseteq \mathcal{L}$ of assumptions, and contrary function $^- : \mathcal{A} \to \mathcal{L}$.*

Unless otherwise specified, we assume a unique contrary a_c for each assumption $a \in \mathcal{A}$. We write $\overline{a}$ to denote a_c.

We fix an arbitrary ABAF $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$. D is *flat* if $head(r) \notin \mathcal{A}$ for any $r \in \mathcal{R}$. A sentence $p \in \mathcal{L}$ is *tree-derivable* via tree t from assumptions $E \subseteq \mathcal{A}$ and rules $R \subseteq \mathcal{R}$, denoted by $E \vdash_{t,R} p$ (tree-derivation), if t is a finite rooted labelled tree such that the root is labelled with p, the set of labels for the leaves of t is equal to E or $E \cup \{\top\}$, and there is a surjective mapping from the set of internal nodes to R, satisfying that for each internal node v, there exists $r \in R$ s.t. v is labelled with $head(r)$, and the set of all successor nodes corresponds to $body(r)$ or $\top$ if $body(r) = \emptyset$. We say p is derivable (in D) iff there is a tree-derivation $E \vdash_{R,t} p$, $E \subseteq \mathcal{A}$, $R \subseteq \mathcal{R}$. We drop R or t if clear from the context.

Definition 3.2. *The derivation $E \vdash p$ is an* ABA argument *if there is a tree-derivation $E \vdash_{R,t} p$ for $E \subseteq \mathcal{A}$, $R \subseteq \mathcal{R}$. $Args_D = \{(A, p) \mid E \vdash_{R,t} p\}$ is the set of all arguments.*

We call arguments of the form $\{a\} \vdash a$, where $a \in \mathcal{A}$, assumption arguments; and arguments with $R \neq \emptyset$ rule-based arguments. We identify assumption arguments with their assumptions. Given an argument $x = E \vdash p$, by $cl(x) = p$ we denote the claim (conclusion) of an argument, by $asms(x) = E$ its assumptions. Given $E \subseteq \mathcal{A}$, we let $\overline{E} = \{\overline{a} \mid a \in E\}$. E is an attacker of $a \in \mathcal{A}$ iff $E \vdash \overline{a}$; E is a supporter of a iff there is $E \vdash_R a$ and $R \neq \emptyset$. Given $a \in \mathcal{A}$, we let $Att(a) = \{E \mid E \vdash \overline{a}\}$ denote the set of all attackers and $Sup(a) = \{E \mid E \vdash a\}$ all supporters. We extend the functions to sets of assumptions: for $E \subseteq \mathcal{A}$, let $Att(E) = \bigcup_{a \in E} Att(a)$; analogously for $Sup(E)$.

3.2 Abstract Representations of ABA

Abstract argumentation (Dung 1995) considers graph-like systems where edges encode relations between abstract entities, typically called arguments (we will also associate them with assumptions). We recall bipolar set-argumentation frameworks (BSAFs) (Berthold, Rapberger, and Ulbricht 2024) that capture (collective) set-attacks and -supports.

Definition 3.3. *A bipolar set-argumentation framework (BSAF) is a tuple $F = (\mathbf{A}, \mathbf{R}, \mathbf{S})$, where $\mathbf{A}$ is a finite set of arguments, $\mathbf{R} \subseteq 2^{\mathbf{A}} \times \mathbf{A}$ is an attack relation and $\mathbf{S} \subseteq 2^{\mathbf{A}} \times \mathbf{A}$ a support relation.*

BSAFs generalise both bipolar AFs (BAFs) (Cayrol and Lagasquie-Schiex 2009) where each $(H, t) \in \mathbf{R} \cup \mathbf{S}$ satisfies $|H| = 1$ and abstract argumentation frameworks (AFs) (Dung 1995) where the same holds but $\mathbf{S} = \emptyset$. For $t \in \mathbf{A}$, we let $Att(t) = \{E \mid (E, t) \in \mathbf{R}\}$ denote the set of all attackers x; then for a set $T \subseteq \mathbf{A}$, we let $Att(T) = \bigcup_{t \in T} Att(t)$; and analogously for $Sup(t)$.

ABAFs are closely related to abstract argumentation (Čyras et al. 2018; Berthold, Rapberger, and Ulbricht 2024). We recall two graph-based representations of ABA. First, we recall the BSAF representation of ABAFs. Here, assumptions are the nodes in the graph and the derivations $S \vdash p$ induce hyperedges that encode attacks and supports.

Definition 3.4. *Let $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-)$ be an ABAF. By $F_D = (\mathcal{A}, \mathbf{R}_D, \mathbf{S}_D)$, we denote the BSAF corresponding to D where $\mathbf{R}_D = Att(\mathcal{A})$ and $\mathbf{S}_D = Sup(\mathcal{A})$ denote the set of all attacks and supports in D, respectively.*

Second, we recall the AF instantiation. Here, the abstract arguments correspond to the ABA arguments and the attacks between the arguments is determined by their claims and assumptions. Note that the instantiation is only defined for *flat* ABAFs where assumptions cannot be derived, i.e., there is no argument $E \vdash_R p$ with $R \neq \emptyset$ and $p \in \mathcal{A}$.

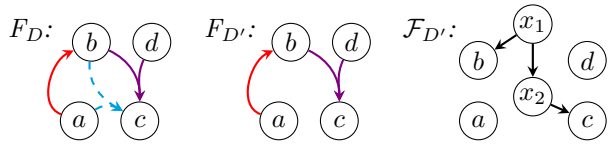
Definition 3.5. *Let $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-)$ be a flat ABAF. By $\mathcal{F}_D = (Args_D, \mathbf{R}_{Args_D})$, we denote the AF corresponding to D where $\mathbf{R}_{Args_D} = \{(x, y) \mid cl(x) \in \overline{asms(y)}\}$.*

Example 3.6. *Consider an ABAF D with assumptions a, b, c, d, their contraries $\overline{a}$, $\overline{b}$, $\overline{c}$, and $\overline{d}$, respectively, and rules*

$$r_1 = c \leftarrow a, b \qquad r_2 = \overline{b} \leftarrow a \qquad r_3 = \overline{c} \leftarrow b, d$$

Note that D is not flat since the assumption c is in the head of r_1. Let D' denote the ABAF without rule r_1.

Below, we depict the BSAF F_D with two attacks $(\{b,d\},c)$ and $(\{a\},b)$ and a support $(\{a,b\},e)$; the BSAF $F_{D'}$; and the AF $\mathcal{F}_{D'}$ with assumption arguments a,b,c,d and two rule-based arguments $x_1 = \{a\} \vdash \overline{b}$ and $x_2 = \{b,d\} \vdash \overline{c}$.



3.3 Gradual Semantics

Quantitative bipolar AFs (QBAFs) (Baroni, Rago, and Toni 2019) generalise AFs and BAFs by introducing an intrinsic strength of arguments.[2]

Definition 3.7. *A quantitative BAF (QBAF) is a tuple $Q = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$ where $(\mathbf{A}, \mathbf{R}, \mathbf{S})$ is a BAF and $\tau : \mathbf{A} \to \mathbb{R}$ assigns each $x \in \mathbf{A}$ a base score.*

Gradual semantics for QBAFs can be defined in terms of an aggregation and influence function, allowing for modular combinations (Mossakowski and Neuhaus 2018). The *aggregation function* has the form $\alpha : \mathbb{R}^{m+k} \mapsto \mathbb{R}$ where the first m variables correspond to the strength values of the attackers and the last k variables to the strengths of the supporters. The order of the attackers and supporters does not play a role; and different arguments may have the same base score. To aid readability, we use multi-sets as input, i.e., sets of real-valued elements that allow for multiple instances for each of its elements, and write $\alpha(A,S)$ for multi-sets A of size m and S of size k. See Table 2 for examples of aggregation functions. The *influence function* has the form $\iota : \mathbb{R}^2 \mapsto \mathbb{R}$. The first variable is the base score of an argument and the second is the aggregate obtained from α, cf. Table 1. All considered aggregation and influence functions in Table 2 and Table 1 respectively satisfy important properties such as *balance* and *monotonicity* (Baroni, Rago, and Toni 2019; Potyka 2019); definitions for which are in Appendix A .

A gradual semantics is the result of the iterative application of a given aggregation and influence function.

Definition 3.8. *Let $Q = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$ be a QBAF with $|\mathbf{A}| = n$, let α be an aggregation and ι be an influence function. The* strength evolution process *is a function $s : \mathbb{R}^n \to \mathbb{R}^n$ defined as follows. For all $a \in \mathbf{A}$, $t \in \mathbb{N}$,*

$$s(0)_a = \tau(a)$$
$$s(t+1)_a = \iota\left(\tau(a), \alpha(A_t^a, S_t^a)\right)$$

where $A_t^a = \{s(t)_b \mid b \in Att(a)\}$ and $S_t^a = \{s(t)_b \mid b \in Sup(a)\}$ denote the strengths of the attackers and supporters of a at time t, respectively.

The modular (α, ι)-semantics σ is defined by the limit

$$\sigma(a) = \lim_{t \to \infty} s(t)_a.$$

Multiple gradual semantics have been studied in the literature, e.g., (Mossakowski and Neuhaus 2018; Amgoud and Ben-Naim 2017). We focus on DF-QuAD (Rago et al. 2016) that uses α_Π and ι^1_{lin} and quadratic energy (QE) (Potyka 2018) that combines α_Σ and ι^1_q.

[2]QBAFs are defined using a set with a preorder in (Baroni, Rago, and Toni 2019) but in this paper we use real numbers $\mathbb{R}$.

$$\iota^k_{lin}(b,w) = b + \tfrac{b}{k}\cdot\min\{0,w\} + \tfrac{1-b}{k}\cdot\max\{0,w\} \quad \textbf{(lin}_k\textbf{)}$$
$$\iota^k_q(b,w) = b - b\cdot h(-\tfrac{w}{k}) + b\cdot h(\tfrac{w}{k}) \quad \textbf{(QE}_k\textbf{)}$$
$$\text{where } h(w) = \tfrac{\max\{0,w\}^2}{(1+\max\{0,w\})^2}$$

Table 1: Influence functions

$$\alpha_\Sigma(A,S) = \Sigma_{s\in S} s - \Sigma_{a\in A} a \quad \textbf{(Sum)}$$
$$\alpha_\Pi(A,S) = \Pi_{a\in A}(1-a) - \Pi_{s\in S}(1-s) \quad \textbf{(Prod)}$$

Table 2: Aggregation functions

4 Gradual ABA Semantics

We introduce gradual semantics for ABA. Thereby, we will utilise BSAFs as they are a syntactical abstraction of ABAFs that concisely represents the interactions between assumptions; all other elements that contribute only indirectly are neglected in the abstraction.

First, let us define quantitative ABAFs.

Definition 4.1. *A quantitative ABAF is a tuple $(\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-, \tau)$ where $(\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-)$ is an ABAF and $\tau : \mathcal{A} \to \mathbb{R}$ assigns a base score to each $a \in \mathcal{A}$.*

In this work, we focus on functions $\tau : \mathcal{A} \to [0,1]$. Throughout the paper, we drop 'quantitative' and simply write ABAF. We introduce *modular* gradual ABA semantics, providing the flexibility to adapt the semantics to different use cases and reuse well-established methods from QBAFs. Our proposed gradual ABA semantics are based on the following procedure, cf. Figure 1. For an ABAF $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-, \tau)$,

1. first, compute the attacks and supports in D to obtain the BSAF representation $F_D = (\mathcal{A}, \mathbf{R}_D, \mathbf{S}_D)$;
2. compute the strength of each set-attack (-support) by applying a *set aggregation function* ζ to the multi-set C corresponding to the attacking (supporting, respectively) set;
3. for each assumption $a \in \mathcal{A}$, use an aggregation function α to update the strength of a, based on the multi-sets A and S corresponding to the strengths of the attackers $Att(a)$ and supporters $Sup(a)$ of a, computed in Step 2;
4. apply the influence function $\iota(b,w)$, with $b = \tau(a)$ the base score of a and $w = \alpha(A,S)$ the result from Step 3,
5. repeat Steps 2-4, using the updated base scores of the assumptions computed in Step 4.

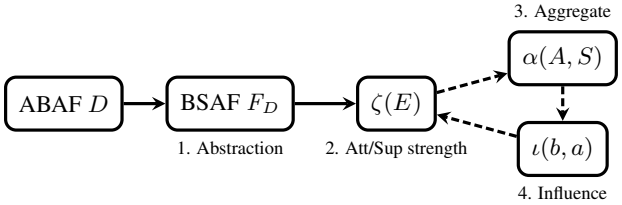


Figure 1: Gradual ABA semantics: Given ABAF D, 1. compute the BSAF F_D, 2. the attacker's and supporter's strength, 3. aggregate, and 4. apply the influence function to update the assumption's strength. Repeat Steps 2 to 4 until convergence (dashed lines).

$\zeta_\Sigma(S) = \Sigma_{s \in S} s$ **(Sum)**	$\zeta_\bot(S) = \min S \cup \{1\}$ **(Min)**
$\zeta_\Pi(S) = \Pi_{s \in S} s$ **(Prod)**	$\zeta_\top(S) = \max S \cup \{0\}$ **(Max)**

Table 3: Set aggregation functions

We will first identify suitable set aggregation functions $\zeta(S)$ in Section 4.1 and discuss their interplay with aggregation functions in Section 4.2. We define gradual ABA semantics and analyse their behaviour in Section 4.3.

4.1 Set Aggregation Functions

We now identify suitable set-aggregation functions to compute the strength of the attacking and supporting sets.

Definition 4.2. *A set aggregation function has the form $\zeta : \mathbb{R}^d \to \mathbb{R}$ whose parameters correspond to the strength values of an (attacking or supporting) set of assumptions. By m_ζ, we denote the maximal value in $\zeta(\mathbb{R}^d)$.*

Since the order of the assumptions should not play a role, we consider multi-sets S as input of ζ and simply write $\zeta(S)$. We stipulate $\zeta(S) \geq 0$ for all S; moreover, for most set aggregation functions that we consider, it holds that $m_\zeta = 1$.

What behaviour do we expect from a set aggregation function ζ? Let us sketch our expectations in an example.

Example 4.3. *Consider an assumption a which is attacked by three sets. For simplicity, we state the attacking sets in terms of their weights: $A_1 = \{0.1, 0.2, 0.9\}$, $A_2 = \{0.9\}$, and $A_3 = \{0, 0.4\}$. We discuss some observations.*

- *The weight of A_2 depends only on a single value. It thus makes intuitively sense to set $\zeta(A_2) = 0.9$.*
- *Comparing A_1 and A_2, we observe that A_2 requires fewer assumptions to attack a, and all weights in A_1 are smaller or equal to the weight in A_2. Intuitively, A_2 should thus be stronger than A_1: the fewer assumptions needed the more effective the attack (similar in spirit to Occam's Razor).*
- *Among the given attacks, A_3 contains the lowest weight (weight 0). Intuitively, an attack should not be stronger than each weakest link (similar in spirit to (Spaans 2021)); thus, A_3 is the weakest attack in this setting.*

Below, we formalise desired principles. For a set aggregation function ζ, we let $S^\circ = \{w \in S \mid w \neq m_\zeta\}$.

Definition 4.4. *A set aggregation function ζ satisfies*

Occam's Razor *iff $S' \subseteq S$ implies $\zeta(S) \leq \zeta(S')$;* ***(OR)***
Factuality *iff $\zeta(\emptyset) = m_\zeta$;* ***(F)***
Neutrality *iff $\zeta(S) = \zeta(S^\circ)$;* ***(N)***
Identity *iff $\zeta(\{s\}) = s$;* ***(ID)***
Weakest link limiting *iff $\zeta(S) \leq \min S \cup \{1\}$;* ***(WL)***
Void *iff $\zeta(S) = 0$ whenever $0 \in S$.* ***(V)***

Void and weakest link limiting appear in the context of argument strength aggregation (Spaans 2021), formalising that an argument cannot be stronger than its weakest link. Identity appears also in (Spaans 2021; Rossit et al. 2021).

We study four basic aggregation methods: sum ζ_Σ, product ζ_Π, minimum $\zeta_\bot$, and maximum $\zeta_\top$. Intuitively, product ζ_Π and minimum $\zeta_\bot$ are more cautious, they resemble conjunctive aggregation methods, while sum ζ_Σ and maximum $\zeta_\top$ follow a more credulous approach by adopting disjunctive aggregation methods. The functions are stated in Table 3. We examine our functions based on the above properties. While ζ_Π and $\zeta_\bot$ satisfy all desired properties, our findings show that ζ_Σ and $\zeta_\top$ are less suited to the purpose.

Proposition 4.5. *ζ_Π and $\zeta_\bot$ satisfy all considered properties; ζ_Σ and $\zeta_\top$ satisfy **(ID)** only.*

We aim to introduce notions of monotonicity and balance for the composition of aggregation and set aggregation functions, similar in spirit to monotonicity and balance for influence and aggregation functions (Potyka 2019; Potyka and Booth 2024). Towards a formalisation, we introduce technical notions that allow us to compare multi-sets.[3] Intuitively, a multi-set A is *superior* to S if A is smaller and wins the pairwise comparison of their elements. In the comparison, we disregard elements with maximal strength, we treat them as facts, similar to Prakken (2018) in the context of probabilistic argumentation. This guarantees that superiority is compatible with our desiderata in Definition 4.4.

Definition 4.6. *Let A, S be two multi-sets. A is superior to S, in symbols $A \succeq_{sup} S$, iff one of the following holds:*

1. *$A^\circ = S^\circ = \emptyset$; or*
2. *for each $S' \subseteq S^\circ$, $|S'| = |A^\circ|$, there is a bijective function $f : A^\circ \to S'$, such that for each $w \in A^\circ$, it holds that $w \geq f(w)$.*

We can simplify the definition slightly, as we show next. Below, we let $Max_k(S)$ denote the multi-set that contains the k largest element of a multi-set S.

Lemma 4.7. *Let A, S be two multi-sets. $A \succeq_{sup} S$ iff $A^\circ = S^\circ = \emptyset$; or there is a bijection $f : A^\circ \to Max_k(S^\circ)$, $k = |A^\circ|$, such that for each $w \in A^\circ$, it holds that $w \geq f(w)$.*

Example 4.8. *Consider $A = \{0.2, 0.2, 0.45, 0.95, 1\}$ and $S = \{0.001, 0.01, 0.2, 0.4, 0.9, 1, 1, 1, 1\}$. The following bijection verifies that A is superior to S, as it maps all elements in $A^\circ = \{0.2, 0.2, 0.45, 0.95\}$ onto elements in $Max_4(S^\circ) = \{0.01, 0.2, 0.4, 0.9\}$ so that $w \geq f(w)$:*

A°	0.95	0.45	0.2	0.2
$f(A^\circ)$	0.9	0.4	0.2	0.01

Next, we introduce *sup-equivalence* and show that superiority and *sup*-equivalence are closely related: if two multi-sets are mutually superior then they are *sup*-equivalent.

Definition 4.9. *Two multi-sets A, S are sup-*equivalent*, in symbols $A \simeq_{sup} S$, iff $A^\circ = S^\circ$.*

Proposition 4.10. *$A \simeq_{sup} S$ iff $A \succeq_{sup} S$ and $S \succeq_{sup} A$.*

4.2 Aggregating Set-Attacks and -Supports

We investigate the interplay between set aggregation and aggregation functions. Thereby, our objects of interest are *multi-sets of multi-sets* corresponding to sets of attackers and supporters, denoted by $\mathbb{A}$ and $\mathbb{S}$, respectively.

[3] A discussion on why dominance, used in (Potyka and Booth 2024), is not suitable here is in Appendix A (Rapberger et al. 2025).

We define dominance and balance below. Intuitively, $\mathbb{A}$ dominates $\mathbb{S}$ if for each $S \in \mathbb{S}$, there is some (distinct) superior set $A \in \mathbb{A}$. Thereby, we ignore sets that contain the number 0 (by the weakest link limiting property, we can disregard them). Below, we let $Pos(\mathbb{A}) = \{A \in \mathbb{A} \mid 0 \notin A\}$.

Definition 4.11. *Let $\mathbb{A}, \mathbb{S}$ be sets of multi-sets.*

Dominance *$\mathbb{A}$ dominates $\mathbb{S}$, in symbols $\mathbb{A} \succeq_{set} \mathbb{S}$, iff* ***(D)***

1. *$Pos(\mathbb{A}) = Pos(\mathbb{S}) = \emptyset$; or*
2. *there is $\mathbb{A}' \subseteq Pos(\mathbb{A})$ and a bijection $f : \mathbb{A}' \to Pos(\mathbb{S})$ s.t. $A \succeq_{sup} f(A)$ for all $A \in \mathbb{A}'$.*

Balance *$\mathbb{A}$ and $\mathbb{S}$ are balanced, in symbols $\mathbb{A} \simeq_{set} \mathbb{S}$, iff $\mathbb{A} \succeq_{set} \mathbb{S}$ and $\mathbb{S} \succeq_{set} \mathbb{A}$.* ***(B)***

In contrast to superiority, dominance requires that the dominating set $\mathbb{A}$ contains at least as many elements as the dominated set $\mathbb{S}$ does (disregarding sets that contain 0).

For a set of sets $\mathbb{A}$, we write $\zeta(\mathbb{A}) = \{\zeta(A) \mid A \in \mathbb{A}\}$.

Definition 4.12. *(Set) aggregation functions α and ζ satisfy (α, ζ)-monotonicity iff*

- $\mathbb{A} \succeq_{set} \mathbb{S} \Rightarrow \alpha(\zeta(\mathbb{A}), \zeta(\mathbb{S})) \leq 0$
- $\mathbb{S} \succeq_{set} \mathbb{A} \Rightarrow \alpha(\zeta(\mathbb{A}), \zeta(\mathbb{S})) \geq 0$
- $\mathbb{A} \succeq_{set} \mathbb{A}' \Rightarrow \alpha(\zeta(\mathbb{A}), \zeta(\mathbb{S})) \leq \alpha(\zeta(\mathbb{A}'), \zeta(\mathbb{S}))$
- $\mathbb{S} \succeq_{set} \mathbb{S}' \Rightarrow \alpha(\zeta(\mathbb{A}), \zeta(\mathbb{S})) \leq \alpha(\zeta(\mathbb{A}), \zeta(\mathbb{S}'))$

Definition 4.13. *A set aggregation function ζ satisfies*

$\succeq$-preservation *iff $\mathbb{A} \succeq_{set} \mathbb{S} \Rightarrow \zeta(\mathbb{A}) \succeq_{sup} \zeta(\mathbb{S})$.* ***($\succeq$-P)***

Proposition 4.14. *ζ_Π and $\zeta_\bot$ satisfy* ***($\succeq$-P)****.*

Note that ζ_Σ and $\zeta_\top$ violate **($\succeq$-P)**. For ζ_Σ, this can be seen, for instance, by taking $A = \{0.2\}$ and $S = \{0.1, 0.1, 0.1\}$. Then $\{A\} \succeq_{set} \{S\}$ since $A \succeq_{sup} S$. However, $\zeta_\Sigma(A) = 0.2$ and $\zeta_\Sigma(S) = 0.3$. We give conditions under which aggregation and set aggregation functions satisfy the combined monotonicity.

Proposition 4.15. *Set aggregation function ζ and aggregation function α satisfy (α, ζ)-monotonicity whenever α satisfies monotonicity and ζ satisfies $\succeq$-preservation.*

4.3 Introducing Gradual ABA Semantics

We are now ready to define and study gradual ABA semantics. As depicted in Figure 1, a key part of the gradual evaluation of an ABAF $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-)$ involves its BSAF representation $F_D = (\mathcal{A}, \mathbf{R}_D, \mathbf{S}_D)$. In this section, we will thus introduce gradual BSAF semantics first and define gradual ABAF semantics based on the BSAF semantics.

Definition 4.16. *A quantitative BSAF is a tuple $F = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$ where $(\mathbf{A}, \mathbf{R}, \mathbf{S})$ is a BSAF and $\tau : \mathbf{A} \to \mathbb{R}$ is a function that assigns a base score to each $a \in \mathbf{A}$.*

In this paper, we focus on functions $\tau : \mathcal{A} \to [0, 1]$. We drop 'quantitative' and simply write BSAF.

Definition 4.17. *Let $F = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$ be a BSAF with $|\mathbf{A}| = n$, let ζ be a set aggregation function, α an aggregation function, and ι an influence function. The* strength evolution process *is a function $s : \mathbb{R}^n \to \mathbb{R}^n$ defined as follows. For all $a \in \mathbf{A}$, $t \in \mathbb{N}$,*

$$s(0)_a = \tau(a)$$
$$s(t+1)_a = \iota\left(\tau(a), \alpha(A^a_t, S^a_t)\right)$$

where $A^a_t = \zeta(\{s(t)_B \mid B \in Att(a)\})$ and $S^a_t = \zeta(\{s(t)_B \mid B \in Sup(a)\})$ are multi-sets containing the aggregated strengths of the attackers and supporters of a at time t, respectively.

The modular (ζ, α, ι)-semantics σ_F is defined by the limit

$$\sigma_F(a) = \lim_{t \to \infty} s(t)_a.$$

We say a semantics σ is well-defined if $\sigma(a)$ converges (admits a unique fixed point) for each $a \in \mathbf{A}$.

Below, we define gradual ABA semantics.

Definition 4.18. *Let $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-, \tau)$ be an ABAF and F_D be the corresponding BSAF. We define $\sigma_D := \sigma_{F_D}$.*

We drop the subscripts F, D, F_D when it is clear from the context. In the remainder of the paper, we identify ABAF and BSAF semantics. All results hold for both semantics.

Remark 4.19. *In the remainder of this section, we associate an ABAF $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-, \tau)$ with the corresponding BSAF $F_D = (\mathcal{A}, \mathbf{R}_D, \mathbf{S}_D, \tau)$. To avoid confusion between ABA arguments and BSAF arguments, we write "assumptions" to refer to the (formally abstract) arguments in F_D.*

The triple (ζ, α, ι) which uniquely defines a modular semantics σ is called the kernel of σ. Below, we adapt the notion of an elementary kernel definition by Potyka (2019).

Definition 4.20. *A kernel is a triple (α, ι, ζ) consisting of an aggregation, influence, and a set-aggregation function. The kernel is elementary iff all functions are Lipschitz-continuous, satisfy monotonicity and balance, α satisfies neutrality and ζ satisfies void and $\succeq$-preservation.*[4]

We study the central properties of gradual semantics (Potyka and Booth 2024) in the context of gradual ABA semantics. Below, we define set-operations component-wise, e.g., $F \cup F' = (\mathbf{A} \cup \mathbf{A}', \mathbf{R} \cup \mathbf{R}', \mathbf{S} \cup \mathbf{S}', \tau \cup \tau')$ for two BSAFs $F = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$, $F' = (\mathbf{A}', \mathbf{R}', \mathbf{S}', \tau')$.

Definition 4.21. *Let $F = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$, $F' = (\mathbf{A}', \mathbf{R}', \mathbf{S}', \tau')$ be two BSAFs. Provided σ_F and $\sigma_{F'}$ are well-defined, a gradual BSAF semantics σ satisfies,*

Anonymity *iff $\sigma_F(a) = \sigma_{F'}(f(a))$ for every edge- and label-preserving graph isomorphism $f : F \to F'$;* ***(A)***

Independence *iff $F \cap F' = \emptyset$ implies $\sigma_{F \cup F'}$ is well-defined and $\sigma_F(a) = \sigma_{F \cup F'}(a)$ for all $a \in A$;* ***(IND)***

Directionality *iff $A = A'$, $\tau = \tau'$, and $A' = A \cup \{(T, h)\}$, then for all $c \in A$, so that there is no directed path from any $t \in T$ to c, $\sigma_F(c) = \sigma_{F'}(c)$.* ***(DIR)***

Individual A-Monotonicity *iff $Att(a) \succeq_{set} Sup(a)$ implies $\sigma(a) \leq \tau(a)$;* ***(IAM)***

[4] Neutrality, monotonicity, and balance for aggregation and influence functions are defined in Appendix A; Lipschitz-continuity is defined in Appendix C (Rapberger et al. 2025).

Individual S-Monotonicity *iff* $Sup(a) \succeq_{set} Att(a)$ *implies* $\sigma(a) \geq \tau(a)$; ***(ISM)***

Relative Monotonicity *iff* $\tau(a) \leq \tau(b)$, $Att(a) \succeq_{set} Att(b)$, *and* $Sup(b) \succeq_{set} Sup(a)$ *implies* $\sigma(a) \leq \sigma(b)$; ***(RM)***

Individual Balance *iff* $Att(a) \simeq_{set} Sup(a)$ *implies* $\sigma(a) = \tau(a)$; ***(IB)***

Relative Balance *iff* $\tau(a) = \tau(b)$, $Att(a) \simeq_{set} Att(b)$, *and* $Sup(b) \simeq_{set} Sup(a)$ *implies* $\sigma(a) = \sigma(b)$; ***(RB)***

We show that anonymity, independence and directionality hold for all modular semantics for BSAFs. For semantics with elementary modular kernel, all remaining properties are satisfied. The theorem below summarises our observations.

Theorem 4.22. *Each modular semantics satisfies* ***(IND)***, ***(A)***, ***(DIR)***. *Each modular semantics with an elementary kernel additionally satisfies* ***(IAM)***, ***(ISM)***, ***(RM)***, ***(IB)***, ***(RB)***.

QBAF semantics are a special case of BSAF semantics, in the following sense.

Proposition 4.23. *Let* $F = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$ *be a BSAF with* $|E| = 1$ *for all* $(E, a) \in \mathbf{R} \cup \mathbf{S}$. *Let* σ *be a modular semantics with kernel* (ζ, α, ι) *where* ζ *satisfies* ***(ID)***. *Then,* σ_F *corresponds to a QBAF semantics with kernel* (α, ι).

We investigate the convergence of modular semantics with elementary kernels. First, if the BSAF is acyclic then convergence is guaranteed. We define cycles with respect to the *primal graph* (Dvorák et al. 2024) where each hyperedge (E, a) is interpreted as simple edges (e, a) for each $e \in E$. A BSAF is acyclic if its corresponding primal graph is.

Proposition 4.24. *Let* $F = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$ *be an acyclic BSAF. Then each elementary modular semantics converges.*

We establish convergence guarantees for the general case. A function f converges with exponential speed to a fixed point x^* if $|f^n(x) - x^*| \leq cd^n$ for some $c \geq 0$ and $0 \leq d < 1$.

Proposition 4.25. *Let* $F = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$ *be a BSAF; let* (ζ, α, ι) *be an elementary kernel with Lipschitz-constants* λ_ζ, λ_α, *and* λ_ι, *respectively. Let* $\lambda_D = \lambda_\zeta \cdot \lambda_\alpha \cdot \lambda_\iota$. *If* $\lambda_D < 1$ *then the strength evolution process* s_D *converges with exponential speed to a fixed point* x^*.

We give convergence guarantees for our proposed semantics. Below, we write $\mathbf{R}(a) = \{(E, a) \mid E \subseteq \mathbf{A}\}$ to denote the set of all attackers of a; $\mathbf{S}(a)$ for supporters.

Proposition 4.26. *Let* $F = (\mathbf{A}, \mathbf{R}, \mathbf{S}, \tau)$ *be a BSAF; let* $d = \max\{|\mathbf{R}(a) \cup \mathbf{S}(a)| \mid a \in \mathbf{A}\}$ *denote the maximum indegree of assumptions in* F *and* $h = \max\{|E| \mid (E, a) \in \mathbf{R} \cup \mathbf{S}\}$ *the maximum size of an attacking or supporting set. Let* α_X *for* $X \in \{\Pi, \Sigma\}$, ι_Y^k *for* $Y \in \{lin, q\}$. *The strength evolution process* s_D *is guaranteed to converge for*

- *for* ζ_Σ, $Z \in \{\Pi, \Sigma\}$, *whenever* $hd < k$;
- *for* ζ_Z, $Z \in \{\top, \bot\}$, *whenever* $d < k$.

For $k = 2$, convergence is guaranteed for $\zeta_\top$ and $\zeta_\bot$ if each element has at most one attacker or supporter. If, in addition, the BSAF corresponds to a QBAF (all attacks and supports have size 1), the result also holds for ζ_Σ and ζ_Π.

Example 4.27. *Let us head back to the ABAF* D *from our introductory Example 1.1. We evaluate* a *wrt.* ζ_Π *and* $\zeta_\bot$ *in combination with the DF-QuAD semantics (*α_Π *and* ι_{lin}^1*).*

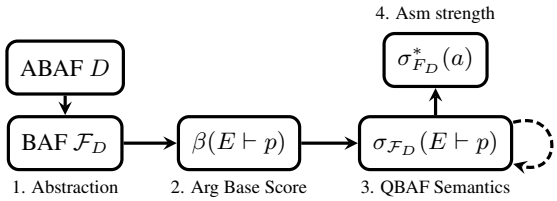


Figure 2: QBAF baseline: Given ABAF D and base score function τ, 1. compute the corresponding BAF $\mathcal{F}_D$, 2. the argument's base score, 3. apply QBAF semantics until convergence (dashed loop), and 4. compute assumption strength.

- *For* τ_1, *the strength of the attack* $(\{b, c\}, a)$ *is* $\zeta_\Pi(\{1, 1\}) = \zeta_\bot(\{1, 1\}) = 1$. *We obtain* $\sigma_D^1(a) = 0$.
- *For* τ_2, *we have* $\zeta_\Pi(\{0.1, 0.2\}) = 0.02$ *and* $\zeta_\bot(\{0.1, 0.2\}) = 0.1$. *Thus,* $(\zeta_\Pi, \alpha_\Pi, \iota_{lin}^1)$ *yields* $\sigma_D^2(a) = 0.98$ *and* $(\zeta_\bot, \alpha_\Pi, \iota_{lin}^1)$ *yields* $\sigma_D^3(a) = 0.9$.

As anticipated, the difference in the base scores is reflected in the final strength. So, if rain (b) and the absence of friends (c) in unlikely, then the strength of a does not decrease much.

5 A QBAF Baseline for ABA

In the realm of extension-based semantics, the direct (assumption-based) semantics for ABA are closely related to the (argument-based) semantics for AFs. It is well-known that the semantics correspond to each other (Bondarenko et al. 1997). We set out to investigate if a similar correspondence can be established for gradual ABA as well. To do so, we use argument base score functions together with the AF representation to instantiate a given ABAF as a QBAF. In this way, we can use QBAF semantics to evaluate an ABAF with weighted assumptions (see Figure 2). Given an ABAF D with base scores τ and a QBAF semantics σ,

1. compute the abstract representation QBAF $\mathcal{F}_D$;
2. compute the argument's base score $\beta(S \vdash p)$, given τ;
3. apply the QBAF semantics and compute $\sigma_{F_D}(S \vdash p)$;
4. compute the strength of the assumptions, $\sigma^*_{F_D}(a)$.

The standard instantiation is defined for flat ABAF only. Below, we extend Definition 3.5 to the non-flat case.

Definition 5.1. *Let* $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-)$ *be an ABAF. By* $\mathcal{F}_D = (Args_D, \mathbf{R}_{Args_D}, \mathbf{S}_{Args_D})$, *we denote the BAF corresponding to* D *where* $\mathbf{R}_{Args_D} = \{(x, y) \mid cl(x) \in \overline{asms(y)}\}$ *and* $\mathbf{S}_{Args_D} = \{(x, y) \mid cl(x) \in asms(y)\}$.

While this instantiation does not preserve extension-based semantics (Ulbricht et al. 2024) it captures the syntactic attack and support relation between the ABA arguments.

We discuss argument base score functions that extend the BAF instantiation to a QBAF in Section 5.1, while methods to extract the assumption strength are in Section 5.2.

5.1 Base Score Function for ABA Arguments

We now identify suitable base score functions. We define a function β that assigns a base score to each argument $S \vdash p$.

Definition 5.2. *Let $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-, \tau)$ be an ABAF. An* argument base score function $\beta : Args_D \to \mathbb{R}$ *assigns a base score to each argument $S \vdash p$, $S \subseteq \mathcal{A}$, $p \in \mathcal{L}$, $|S| = d$.*

We consider *syntax independence* (called anonymity in (Spaans 2021) and related to rewriting in (Jedwabny, Croitoru, and Bisquert 2020)) which intuitively means that the base score of an argument is independent of its structure.

Definition 5.3. *An argument base score function β satisfies* **Syntax Independence** *iff $\beta(S \vdash_t p) = \beta(S \vdash_{t'} p)$.* ***(SI)***

We consider a class of argument base score functions that neglect the specific argument tree; the strength is entirely determined by the assumptions in the argument.

Definition 5.4. *An assumption-based argument base score function $\beta : \mathbb{R}^d \to \mathbb{R}$ assigns each argument $S \vdash_t p$, $|S| = d$, a base score based on the assumption's strengths.*

We focus on assumption-based functions. This choice has convenient implications: each such argument base score function satisfies syntax independence (**SI**) by design. Most of the desirable properties from Definition 4.4 in the context of set aggregation functions apply to the case of argument base score functions as well. Occam's razor, for instance, formalises that the fewer assumptions the stronger an argument. As shown in Section 4.1, the product and minimum function satisfy all desired properties. We will thus consider the following argument base score functions.

$\beta_\Pi(E) = \prod_{e \in E} e$ **(Prod)** $\quad \beta_\perp(E) = \min E \cup \{1\}$ **(Min)**

Both functions satisfy syntax independence by definition.

Proposition 5.5. *β_Π and $\beta_\perp$ satisfy **(SI)**, **(OR)**, **(F)**, **(N)**, **(ID)**, **(WL)**, and **(V)**.*

5.2 Assumption Strength Computation

We investigate suitable functions to determine the final strength of the assumptions, based on the final strength of the arguments that derive them. That is, to compute the strength of a, we take all arguments $S \vdash a$ into account. Since a single claim may be supported by several arguments, we have several choices to obtain the strength of the assumptions.

Definition 5.6. *Let $D = (\mathcal{L}, \mathcal{A}, \mathcal{R}, ^-, \tau)$ be a ABAF, σ be a QBAF semantics, and let $S \subseteq Args_D$ denote the set that contains all arguments with claim a, i.e., $S = \{E \vdash a \mid E \subseteq \mathcal{A}\}$, let $x = \{a\} \vdash a$ for each $a \in \mathcal{A}$. We define*

$$\sigma^*_{\text{asm}}(S) = \sigma(x_a) \textbf{ (Asm)} \qquad \sigma^*_\perp(S) = \min S \textbf{ (Min)}$$
$$\sigma^*_{\text{avg}}(S) = \tfrac{1}{|S|} \sum S \textbf{ (Avg)} \qquad \sigma^*_\top(S) = \max S \textbf{ (Max)}$$

Note σ^*_{avg} averages the outcome while the other functions represent choices to pick a representative strength. For flat ABAFs, all of the options coincide (for assumptions) since we have only one argument for each assumption.

Proposition 5.7. *Let D be a flat ABAF and σ a QBAF semantics. Then $\sigma = \sigma'$ for all $\sigma, \sigma' \in \{\sigma^*_{\text{asm}}, \sigma^*_\perp, \sigma^*_\top, \sigma^*_{\text{avg}}\}$.*

We note that convergence depends on the choice of the QBAF semantics, i.e., σ^* converges iff σ does.

We investigate the relation of the QBAF baseline to gradual ABA semantics. Note that the topology of a BSAF representation will typically differ from the BAF instantiation (see, for instance, Example 3.6). We obtain equality in the special case in which no rules are given.

Proposition 5.8. *Let $D = (\mathcal{L}, \mathcal{A}, \emptyset, ^-, \tau)$ be an ABAF, F_D denote the BSAF corresponding to D, $\mathcal{F}_D$ denote the QBAF corresponding to D, σ_D be a modular (ζ, α, ι)-semantics for D and $\sigma_{\mathcal{F}_D}$ denote the corresponding modular (α, ι)-semantics. It holds that $\sigma_D = \sigma_{F_D} = \sigma_{\mathcal{F}_D}$.*

6 Experimental Evaluation

We now explore the differences between our proposed gradual ABA (BSAF) and the presented QBAF baseline (BAF) and within various modules of the semantics. Our experiments show that BSAFs converge 90% of the scenarios while BAFs do so only in around 70% of the scenarios; moreover, BSAFs converge faster (in around 30 iterations) than BAFs do (in around 45 iterations).

Experimental Setup. We evaluate BSAF and BAF on a suite of 1440 randomly generated ABAFs (960 non-flat, 480 flat), following the procedure in (Lehtonen et al. 2024). Details are provided in Appendix E.1 (Rapberger et al. 2025).

The runtime of both BSAF and BAF approaches depends on the instantiations (ABAF to SETAF and ABAF to AF, respectively), which can be both exponential in the number of assumptions in the worst case, and on the number of passes, which depends on the structure of the resulting abstract instance. As discussed in the previous sections, convergence cannot always be guaranteed; therefore, we constrained the number of iterations. Each run has a 10-minute time budget and up to 5000 iterations; we record 92 timeouts during abstraction (all on non-flat ABAFs), with 10 additional timeouts for BSAF and 89 for BAF in the semantics phases (see Figures 1 and 2).

Convergence Metrics. We use two metrics to quantify the convergence behaviour of our gradual ABA semantics:

- *Global convergence rate:* proportion of ABAFs in which every assumption strength settles (within $\epsilon = 10^{-3}$) over the final $\delta = 5$ updates before hitting the cap on the number of iterations.
- *Average steps to converge:* mean iteration step to convergence, computed only over the runs that do converge.

Further methodological details are in Appendix E.2 (Rapberger et al. 2025).

Modular Components Combinations. We test two types for each of the modules presented in the previous sections:

- Aggregations & Influence Functions:
  - *DF-QuAD* (DFQ): α_Π with linear influence ι^1_{lin}
  - *Quadratic Energy* (QE): α_Σ with quadratic influence ι^1_q
- Sets Aggregation & Base-score Functions:
  - *BSAF:* Att/Sup aggregation $\zeta \in \{\zeta_\perp, \zeta_\Pi\}$.
  - *BAF:* Base score aggregation $\beta \in \{\beta_\perp, \beta_\Pi\}$.

We use two base score initialisation: constant ($\tau(a) = 0.5$) and random ($\tau(a) \sim U(0,1)$) and two assumptions strengths functions $\sigma^* \in \{\sigma^*_{\text{avg}}, \sigma^*_{\text{asm}}\}$ (for BAF only). Results for the latter ablation, with no significant differences observed, are in Appendix E.3 (Rapberger et al. 2025).

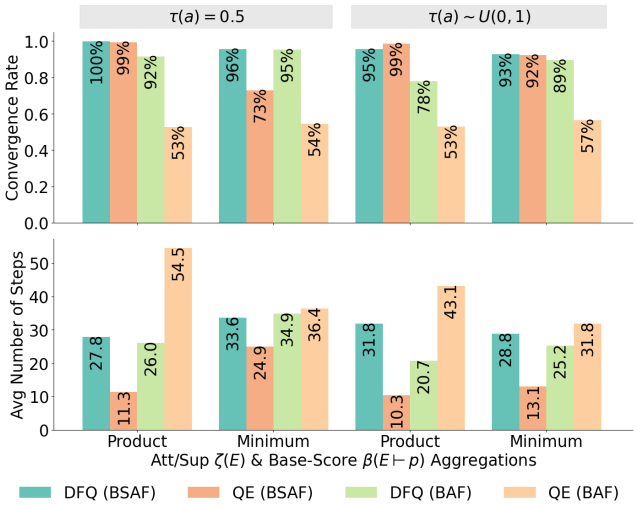


Figure 3: Global convergence rates (top, y-axis) and average steps to converge (bottom, y-axis) for the BSAF and BAF approach. We compare two modular semantics (DFQ vs. QE), two Att/Sup and Base-score aggregations (x-axis, resp. for BSAF and BAF), across two base-score initialisations ($\tau = 0.5$ vs. $\tau \sim U(0, 1)$).

BSAF vs. BAF. We show global convergence rates (top) and average steps to converge (bottom) in Figure 3. Overall:

1. **Convergence Rate:** BSAF achieves convergence in the range of 73-95% of runs, compared to 53-95% under BAF, with DFQ leading across both BSAFs and BAFs.
2. **Convergence Speed:** BSAF converges in 10-30 iterations, while BAF requires 20-55.
3. **Sensitivity to Initialisation:** BSAF is insensitive to τ, while BAF shows wide variance across initialisations.

DFQ vs. QE within BSAF. DFQ consistently achieves $\geq$90% convergence across all settings. QE converges fastest under Product aggregation (circa 10 steps for both $\tau = 0.5$ and $\tau \sim U(0, 1)$) but slows to 25-35 steps under Minimum. DFQ suffers from random initialisation (vs. constant), while QE benefits, possibly reflecting the open-mindedness vs. conservativeness trade-off in (Potyka and Booth 2024).

To explore the differences between DFQ and QE in more detail, we conduct an additional analysis on how convergence rates vary by structural ABAFs' features. The summary of the results is shown in Figure 4. We split convergence rates by: *number of assumptions* (left column); *number of rules* (center column); *average rule body size* (right column). Results are given for constant ($\tau = 0.5$, top row) and random ($\tau \sim U(0, 1)$, bottom row) base scores initialisations. DFQ remains above 90% in every bucket, whereas QE degrades on larger, more complex instances. More comparisons, breakdowns and examples of strengths' trajectories are in Appendix E.3 (Rapberger et al. 2025).

Overall, BSAF outperforms BAF in both convergence reliability and speed. Within BSAF, DFQ provides the greatest robustness across ABAFs' characteristics and initialisations, while QE trades some resilience for rapid convergence.

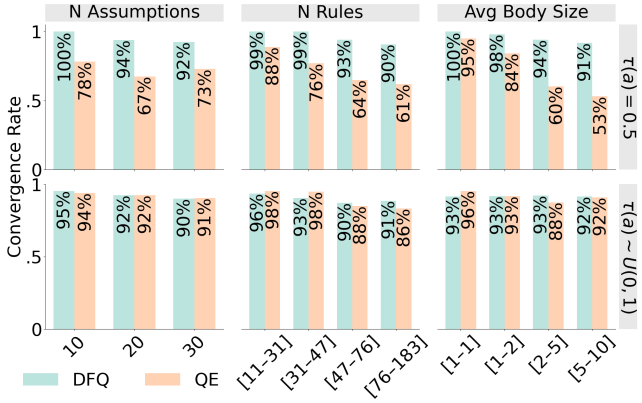


Figure 4: Sensitivity of the gradual ABA semantics (BSAF) to ABAF structural characteristics. Convergence rates are binned by number of assumptions (left), number of rules (center), and average rule body size (right), for both DFQ and QE under constant ($\tau = 0.5$, top row) and random ($\tau \sim U(0, 1)$, bottom row) base-scores. The set aggregation is minimum (product in Appendix E.3).

7 Conclusion

We have presented a family of novel gradual semantics for ABA, filling a gap in the literature and bridging to the extensive body of work on gradual semantics for abstract forms of argumentation. Our gradual semantics is defined via BSAFs that allow us to focus on the (supporting and attacking) relations between the assumptions. We have formally and experimentally evaluated our gradual ABA semantics against a QBAF-based baseline for ABA and have shown that it exhibits desirable characteristics.

Future work. By presenting the very first purpose-built gradual semantics for ABA, our work opens several avenues for future work. One regards the correspondence of gradual ABA semantics to other evaluation paradigms such as extension-based or labelling-based semantics. In this aspect, it would be interesting to investigate gradual interpretations of the *closure property* known from extension-based semantics as the requirement that ensures that an extension contains all assumptions that it derives. Also, extending gradual semantics to ABA with preferences using set-to-set attacks (Dimopoulos et al. 2024) would be an interesting avenue for future research. In this work, we assume that the non-assumptions of our knowledge base (i.e., rules and sentences) do not have an intrinsic strength on their own.

For future work, we want to investigate settings with varying intrinsic strength of sentences that are non-assumptions; and settings in which the strength of rules is part of the evaluation. Given that assumptions can render the rules defeasible, we can introduce rule weights using a dedicated assumption for each rule which represents its weight. In this way, the assumptions' weight can be equated to the rules' weights. Our argument base score functions could also be used to evaluate the strength of ordinary conclusions. We also plan to explore concrete applications of our gradual semantics, e.g., for causal discovery with the ABA framework in (Russo, Rapberger, and Toni 2024).

Acknowledgments

Rapberger and Russo were funded by the European Research Council (ERC) under the ERC-POC programme (grant agreement no. 101189053). Rago and Toni were partially funded by the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 101020934) and by J.P. Morgan and the Royal Academy of Engineering under the Research Chairs and Senior Research Fellowships scheme (grant agreement no. RCSRF2021\11\45).

References

Amgoud, L., and Ben-Naim, J. 2015. Argumentation-based ranking logics. In *AAMAS*, 1511–1519.

Amgoud, L., and Ben-Naim, J. 2017. Evaluation of arguments in weighted bipolar graphs. In *ECSQARU*, 25–35.

Amgoud, L., and Doder, D. 2018. Gradual semantics for weighted graphs: An unifying approach. In *KR*, 613–614.

Amgoud, L.; Cayrol, C.; Lagasquie-Schiex, M.; and Livet, P. 2008. On bipolarity in argumentation frameworks. *Int. J. Intell. Syst.* 23(10):1062–1093.

Ayoobi, H.; Potyka, N.; and Toni, F. 2023. Sparx: Sparse argumentative explanations for neural networks. In *ECAI*, 149–156.

Baroni, P.; Caminada, M.; and Giacomin, M. 2018. Abstract argumentation frameworks and their semantics. In *Handb. Formal Argument.* College Publications. 159–236.

Baroni, P.; Rago, A.; and Toni, F. 2019. From fine-grained properties to broad principles for gradual argumentation: A principled spectrum. *Int. J. Approx. Reason.* 105:252–286.

Berthold, M.; Rapberger, A.; and Ulbricht, M. 2023. On the expressive power of assumption-based argumentation. In *JELIA*, 145–160.

Berthold, M.; Rapberger, A.; and Ulbricht, M. 2024. Capturing non-flat assumption-based argumentation with bipolar SETAFs. In *KR*, 128–133.

Besnard, P., and Hunter, A. 2014. Constructing argument graphs with deductive arguments: a tutorial. *Argument Comput.* 5(1):5–30.

Bondarenko, A.; Dung, P. M.; Kowalski, R. A.; and Toni, F. 1997. An abstract, argumentation-theoretic approach to default reasoning. *Artif. Intell.* 93:63–101.

Caminada, M., and Gabbay, D. M. 2009. A logical account of formal argumentation. *Stud. Log.* 93(2-3):109–145.

Caminada, M., and Schulz, C. 2017. On the equivalence between assumption-based argumentation and logic programming. *J. Artif. Intell. Res.* 60:779–825.

Cayrol, C., and Lagasquie-Schiex, M. 2005. Graduality in argumentation. *J. Artif. Intell. Res.* 23:245–297.

Cayrol, C., and Lagasquie-Schiex, M. 2009. Bipolar abstract argumentation systems. In *Argumentation in Artificial Intelligence*. Springer. 65–84.

Čyras, K., and Toni, F. 2016. ABA+: assumption-based argumentation with preferences. In *KR*, 553–556.

Čyras, K.; Fan, X.; Schulz, C.; and Toni, F. 2018. Assumption-based argumentation: Disputes, explanations, preferences. In *Handb. Formal Argument.* College Publications. 365–408.

Čyras, K.; Oliveira, T.; Karamlou, A.; and Toni, F. 2021. Assumption-based argumentation with preferences and goals for patient-centric reasoning with interacting clinical guidelines. *Argument Comput.* 12(2):149–189.

Čyras, K.; Heinrich, Q.; and Toni, F. 2021. Computational complexity of flat and generic assumption-based argumentation, with and without probabilities. *Artif. Intell.* 293:103449.

Dimopoulos, Y.; Dvorák, W.; König, M.; Rapberger, A.; Ulbricht, M.; and Woltran, S. 2024. Redefining ABA+ semantics via abstract set-to-set attacks. In *AAAI*, 10493–10500.

Dung, P. M., and Thang, P. M. 2010. Towards (probabilistic) argumentation for jury-based dispute resolution. In *COMMA*, 171–182.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77(2):321–357.

Dvorák, W.; König, M.; Ulbricht, M.; and Woltran, S. 2024. Principles and their computational consequences for argumentation frameworks with collective attacks. *J. Artif. Intell. Res.* 79:69–136.

Fan, X. 2023. Probabilistic deduction as a probabilistic extension of assumption-based argumentation. In *AAMAS*, 2352–2354.

Gonzalez, M. G. E.; Budán, M. C. D.; Simari, G. I.; and Simari, G. R. 2021. Labeled bipolar argumentation frameworks. *J. Artif. Intell. Res.* 70:1557–1636.

Hecham, A.; Bisquert, P.; and Croitoru, M. 2018. On a flexible representation for defeasible reasoning variants. In *AAMAS*, 1123–1131.

Heyninck, J.; Raddaoui, B.; and Straßer, C. 2023. Ranking-based argumentation semantics applied to logical argumentation. In *IJCAI*, 3268–3276.

Hunter, A. 2013. A probabilistic approach to modelling uncertain logical arguments. *Int. J. Approx. Reason.* 54(1):47–81.

Hunter, A. 2022. Argument strength in probabilistic argumentation based on defeasible rules. *Int. J. Approx. Reason.* 146:79–105.

Jedwabny, M.; Croitoru, M.; and Bisquert, P. 2020. Gradual semantics for logic-based bipolar graphs using t-(co)norms. In *ECAI*, 777–783.

Lehtonen, T.; Rapberger, A.; Ulbricht, M.; and Wallner, J. P. 2023. Argumentation frameworks induced by assumption-based argumentation: Relating size and complexity. In *KR*, 440–450.

Lehtonen, T.; Rapberger, A.; Toni, F.; Ulbricht, M.; and Wallner, J. P. 2024. Instantiations and computational aspects of non-flat assumption-based argumentation. In *IJCAI*, 3457–3465.

Lehtonen, T.; Wallner, J. P.; and Järvisalo, M. 2021a. Declarative algorithms and complexity results for assumption-based argumentation. *J. Artif. Intell. Res.* 71:265–318.

Lehtonen, T.; Wallner, J. P.; and Järvisalo, M. 2021b. Harnessing incremental answer set solving for reasoning in assumption-based argumentation. *Theory Pract. Log. Program.* 21(6):717–734.

Leite, J., and Martins, J. G. 2011. Social abstract argumentation. In *IJCAI*, 2287–2292.

Modgil, S., and Prakken, H. 2014. The *ASPIC*$^+$ framework for structured argumentation: a tutorial. *Argument Comput.* 5(1):31–62.

Mossakowski, T., and Neuhaus, F. 2018. Modular semantics and characteristics for bipolar weighted argumentation graphs. *CoRR* abs/1807.06685.

Oluokun, B.; Paulino-Passos, G.; Rago, A.; and Toni, F. 2024. Predicting human judgement in online debates with argumentation. In *CMNA 2024*, 70–80.

Potyka, N., and Booth, R. 2024. Balancing Open-Mindedness and Conservativeness in Quantitative Bipolar Argumentation (and How to Prove Semantical from Functional Properties). In *KR*, 597–607.

Potyka, N. 2018. Continuous dynamical systems for weighted bipolar argumentation. In *KR*, 148–157.

Potyka, N. 2019. Extending modular semantics for bipolar weighted argumentation. In *AAMAS*, 1722–1730.

Prakken, H. 2018. Probabilistic strength of arguments with structure. In *KR*, 158–167.

Prakken, H. 2024. An abstract and structured account of dialectical argument strength. *Artif. Intell.* 335:104193.

Rago, A.; Toni, F.; Aurisicchio, M.; and Baroni, P. 2016. Discontinuity-free decision support with quantitative argumentation debates. In *KR*, 63–73.

Rago, A.; Vasileiou, S. L.; Toni, F.; Son, T. C.; and Yeoh, W. 2024. A methodology for incompleteness-tolerant and modular gradual semantics for argumentative statement graphs. *CoRR* abs/2410.22209.

Rago, A.; Cocarascu, O.; Oksanen, J.; and Toni, F. 2025. Argumentative review aggregation and dialogical explanations. *Artif. Intell.* 340:104291.

Rapberger, A.; Russo, F.; Rago, A.; and Toni, F. 2025. On gradual semantics for assumption-based argumentation. *CoRR* abs/2507.10076.

Rossit, J.; Mailly, J.; Dimopoulos, Y.; and Moraitis, P. 2021. United we stand: Accruals in strength-based argumentation. *Argument Comput.* 12(1):87–113.

Russo, F.; Rapberger, A.; and Toni, F. 2024. Argumentative causal discovery. In *KR*, 938–949.

Schulz, C., and Toni, F. 2017. Labellings for assumption-based and abstract argumentation. *Int. J. Approx. Reason.* 84:110–149.

Skiba, K.; Thimm, M.; and Wallner, J. P. 2023. Ranking-based semantics for assumption-based argumentation. In *FCR*, 44–52.

Skiba, K.; Thimm, M.; and Wallner, J. P. 2024. Ranking transition-based medical recommendations using assumption-based argumentation. In *RATIO*, 202–220.

Spaans, J. P. 2021. Intrinsic argument strength in structured argumentation: A principled approach. In *CLAR*, 377–396.

Toni, F. 2013. A generalised framework for dispute derivations in assumption-based argumentation. *Artif. Intell.* 195:1–43.

Ulbricht, M.; Potyka, N.; Rapberger, A.; and Toni, F. 2024. Non-flat ABA is an instance of bipolar argumentation. In *AAAI*, 10723–10731.

Zeng, Z.; Shen, Z.; Chin, J. J.; Leung, C.; Wang, Y.; Chi, Y.; and Miao, C. 2020. Explainable and contextual preferences based decision making with assumption-based argumentation for diagnostics and prognostics of alzheimer's disease. In *AAMAS*, 2071–2073.