Bounding Acceptability Degrees and Eliciting Initial Weights in Gradual Argumentation

Nir Oren[1] , **Bruno Yun**[2]
[1]University of Aberdeen
[2]Universite Claude Bernard Lyon 1, CNRS, Ecole Centrale de Lyon, INSA Lyon
n.oren@abdn.ac.uk, bruno.yun@univ-lyon1.fr

Abstract

Many semantics for abstract weighted argumentation assume that each argument is associated with a numerical initial weight. Eliciting these initial weights poses several challenges: (1) accurately providing a specific numerical value is often difficult, and (2) individuals frequently confuse initial weights with acceptability degrees in the presence of other arguments. We therefore propose an elicitation pipeline that allows a user to specify their believed final acceptability degree intervals for each argument. We can determine which portion (if any) of these intervals are rational, refining the intervals, or restoring rationality when the intervals are irrational. This allows us to ultimately identify possible initial weights for each argument.

1 Introduction

After the seminal work of [Dung, 1995], the argumentation community's focus has shifted to the relation between arguments, abstracting the content of the arguments. Debates or discussions are represented as directed graphs, with arguments as nodes and attacks between arguments as arcs. Several *extension-based semantics* have been defined to obtain conclusions from those graphs. Such semantics identify subsets of arguments (called extensions), representing consistent conclusions [Baroni *et al.*, 2011]

Motivated by the work of [Amgoud and Ben-Naim, 2013a], researchers started focusing on semantics which could give a more gradual view on arguments' acceptability by ranking them from the least acceptable to the most acceptable (not necessarily based only on the cardinality or quality of the attackers) [Bonzon *et al.*, 2018]. A subset of these semantics, called *gradual semantics*, assigns to each argument a numeric (final) acceptability degree representing its strength after taking into account the relation between arguments (e.g., [Besnard and Hunter, 2001]). More recent works expand gradual semantics with supports [Rago *et al.*, 2016], sets of attacking arguments [Yun *et al.*, 2020], and initial weights [Amgoud and Ben-Naim, 2018].

At the same time, it is becoming increasingly recognised that people struggle to differentiate between an argument's initial weight and its final acceptability degree. For example, in the context of Dung-style arguments, [Rahwan *et al.*, 2010] found that individuals assign lower strength to reinstated arguments, suggesting that individuals report acceptability degree instead of initial weights when reasoning with an argumentation graph (as initial weights should be unchanged). Moreover, [Vesic *et al.*, 2022] showed that individuals reason better when presented with an argumentation graph (as opposed to only textual arguments). We thus argue that when eliciting information about arguments, it may make more sense to obtain what individuals perceive as an argument's final acceptability degree *in the context of the entire argumentation system*, rather than try to obtain its initial weights.

Moreover, humans are rarely precise when specifying argument strengths. Rather, they use natural language terms (e.g., "this argument is convincing", or "this argument is stronger than that argument") or utilise ranges (e.g., "likely" may indicate a 66-90% confidence [IPCC, 2023, pg.3]).

Building on these observations, we make two primary contributions. First, we investigate how to narrow down an argumentation system's final acceptability degrees for each argument given an interval over final acceptability degrees, based on some desirable properties, using the given values as well as the underlying argumentation framework structure. We term the resultant intervals *rational*. If no such rational intervals exist, we seek to find rational values which require minimal changes from elicited values. We also provide an implementation and evaluation of our system[1]. Our second contribution involves identifying the range of initial weights obtainable from a final acceptability degree interval, recognising that the mapping between the two is non-linear.

2 Background

A weighted argumentation framework is composed of arguments, attacks between arguments, and a weighting function associating each argument to an initial weight representing its trustworthiness or certainty degree of its premises.

Definition 1. *A weighted argumentation framework (WAF) is* $\mathbf{WAF} = (\mathcal{A}, \mathcal{R}, w)$*, where* $\mathcal{A}$ *is a finite set of arguments,* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is a set of attacks between arguments, and* w *is a weighting function from* $\mathcal{A}$ *to* $[0, 1]$.

[1]A user interface demonstrating the system (with the link to source code) is available at https://weight-elicitator.vercel.app/.

Given $\mathbf{wAF} = (\mathcal{A}, \mathcal{R}, w)$ and $a \in \mathcal{A}$, $Att(a) = \{b \in \mathcal{A} \mid (b, a) \in \mathcal{R}\}$ and $Att^*(a) = \{b \in Att(a) \mid w(a) \neq 0\}$.

Dung's extension-based semantics [Dung, 1995] induce a two-level acceptability of arguments (inside or outside of one or all extensions). Gradual semantics (and ranking-based semantics in general) have been proposed as a more fine-grained approach to argument acceptability [Amgoud and Ben-Naim, 2013b; Bonzon *et al.*, 2023], and their weighted versions have been defined (e.g., [Amgoud *et al.*, 2022]). These gradual semantics use the weighting to compute a score for each argument in the weighted argumentation framework, called its (acceptability) degree.

Definition 2. *A weighted gradual semantics is a function σ which associates to each $\mathbf{wAF} = (\mathcal{A}, \mathcal{R}, w)$, a weighting $\sigma_{\mathbf{wAF}} : \mathcal{A} \to [0, 1]$ on $\mathcal{A}$. $\sigma_{\mathbf{wAF}}(a)$ is the degree of a.*

We recall the well-known weighted gradual semantics studied by [Besnard and Hunter, 2001; Oren *et al.*, 2022b; Amgoud *et al.*, 2022]. The weighted h-categoriser semantics (`Hbs`) assigns a value to each argument by considering the initial weight of the argument and degree of its attackers, which themselves take into account the degree of their attackers.

Definition 3. *The weighted h-categoriser semantics is a gradual semantics $\sigma^{\texttt{Hbs}}$ s.t. for any $\mathbf{wAF} = (\mathcal{A}, \mathcal{R}, w)$ and $a \in \mathcal{A}$, $\sigma^{\texttt{Hbs}}_{\mathbf{wAF}}(a) = w(a)/(1 + \Sigma_{b \in Att(a)} \sigma^{\texttt{Hbs}}_{\mathbf{wAF}}(b))$.*

Weighted Card-based semantics (`Car`) favour the number of attackers over their quality. This semantics is based on a recursive function which assigns a score to arguments on the basis initial weights, number of attackers, and their degrees.

Definition 4. *The weighted Card-based semantics is a gradual semantics $\sigma^{\texttt{Car}}$ s.t. for any $\mathbf{wAF} = (\mathcal{A}, \mathcal{R}, w)$ and $a \in \mathcal{A}$, $\sigma^{\texttt{Car}}_{\mathbf{wAF}}(a) = w(a)/(1 + |Att^*(a)| + \frac{\Sigma_{b \in Att^*(a)} \sigma^{\texttt{Car}}_{\mathbf{wAF}}(b)}{|Att^*(a)|})$ if $Att^*(a) \neq \emptyset$, otherwise $\sigma^{\texttt{Car}}_{\mathbf{wAF}}(a) = w(a)$.*

Weighted Max-based semantics (`Max`) favour the quality of attackers over their number. An argument's degree is based on its initial weight and degree of its strongest direct attacker.

Definition 5. *The weighted Max-based semantics is a gradual semantics $\sigma^{\texttt{Max}}$ s.t. any $\mathbf{wAF} = (\mathcal{A}, \mathcal{R}, w)$ and $a \in \mathcal{A}$, $\sigma^{\texttt{Max}}_{\mathbf{wAF}}(a) = w(a)/(1 + \max_{b \in Att(a)} \sigma^{\texttt{Max}}_{\mathbf{wAF}}(b))$*

Starting from a weighted argumentation framework, one utilises a weighted gradual semantics to compute acceptability degrees. The approach of [Oren *et al.*, 2022b] has shown that we can instead start with acceptability degrees and the graph to infer initial weights. In this work, we explore what to do *if we are unsure about the acceptability degrees.*

3 Working with Intervals

Humans often struggle to differentiate between an argument's initial weights and its acceptability degree, as human reasoning implicitly considers additional background knowledge. Moreover, humans also struggle to pinpoint specific acceptability degree values. We thus propose a framework where one can indicate an interval for acceptability degrees, from which appropriate final acceptability degrees can be identified and possible initial weights can be sampled.

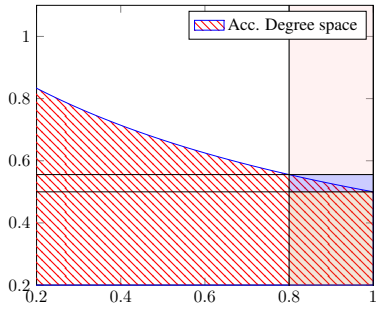


Figure 1: The hatched area represents $D^{\mathbf{AF}}_{\sigma}$. The x- and y-axes are the degree of a and b. Vertical lines are at $x_1 = 0.8$ and $x_2 = 1$. Horizontal lines are at $y_1 = 0.5$ and $y_2 = 5/9$.

Namely, the process is as follows. Individuals would first provide intervals (step 1). If the input provided is rational (w.r.t. some weighted gradual semantics), then we refine the interval of linked arguments using the corresponding gradual semantics (step 2a). However, in the case where the appraisal is irrational (for a given gradual semantics), the user will have to re-do their appraisal with support from the framework (step 2b). Note that after the automatic refinement of the intervals, the user can continue to improve their input. Once users are satisfied with their acceptability degree intervals, they can sample some possible valid initial weights.

Definition 6. *An interval constrained argumentation framework (ICAF) is $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$, where $\mathbf{AF} = (\mathcal{A}, \mathcal{R})$ is an (non-weighted) argumentation framework and I is a function s.t. $\forall a \in \mathcal{A}, I(a) \subseteq [0, 1]$.*

In our setting, an interval $I(a) = [0, 1]$ for some argument a effectively means that there is no restriction on a.

Definition 7. *Given $\mathbf{AF} = (\mathcal{A}, \mathcal{R})$ and a weighted gradual semantics σ, where $\mathcal{A} = \{a_1, \ldots, a_n\}$ and a point $P = (p_1, \ldots, p_n) \in [0, 1]^n$. We say that P is in the* acceptability degree space *w.r.t. σ iff $\exists w$ s.t. $\forall a_i \in \mathcal{A}, \sigma_{(\mathcal{A},\mathcal{R},w)}(a_i) = p_i$.*

We call the acceptability degree space (of $\mathbf{AF}$ w.r.t. σ), denoted $D^{\mathbf{AF}}_{\sigma}$, the set of all such points [Oren et al.*, 2022b].*

Definition 8. *Let σ be a weighted gradual semantics and $\mathbf{ICAF} = (\mathcal{A} = \{a_1, a_2, \ldots, a_n\}, \mathcal{R}, I)$ be an ICAF. A pair $(\mathbf{ICAF}, \sigma)$ is* irrational *if $\nexists w : \mathcal{A} \to [0, 1]$ s.t. $\forall a_i \in \mathcal{A}$, $\sigma_{(\mathcal{A},\mathcal{R},w)}(a_i) \in I(a_i)$, and is* rational *otherwise. $(\mathbf{ICAF}, \sigma)$ is fully rational if $\forall (d_1, d_2, \ldots, d_n)$, where $d_i \in I(a_i)$ for all $1 \leq i \leq n$, $\exists w : \mathcal{A} \to [0, 1]$ s.t. $\sigma_{(\mathcal{A},\mathcal{R},w)(a_i)} = d_i$.*

Example 1. *Consider an ICAF $(\{a, b\}, \{(a, b)\}, I)$. In Figure 1, we represent the acceptability degree space of a and b (below the blue curve) w.r.t. the weighted h-categoriser. For every point (d_a, d_b) in this area, there exists a weighting function w s.t. $\sigma_{(\mathcal{A},\mathcal{R},w)}(a) = d_a$ and $\sigma_{(\mathcal{A},\mathcal{R},w)}(b) = d_b$.*

Assume that $I(a) = [0.8, 1.0]$. Let $U = 5/9$ and $L = 0.5$.

- *If $I(b) \in [x, L]$, for $0 \leq x \leq L$, then the ICAF is fully rational as for every (d_a, d_b), $d_a \in I(a)$, and $d_b \in I(b)$, there exists a weighting function w s.t. $\sigma_{(\mathcal{A},\mathcal{R},w)}(a) = d_a$ and $\sigma_{(\mathcal{A},\mathcal{R},w)}(b) = d_b$. Any such point (d_a, d_b) will lie within the brown rectangle shown in Figure 1. Thus, any point in $I(a) \times I(b)$ lies within the rectangle.*

- *If $I(b) = [x, y]$, for $L < y, 0 \leq x \leq U$ and $x \leq y$, then the ICAF is rational, as the point $(0.8, x) \in I(a) \times I(b)$ is in the brown or blue area but it is not fully rational as the point $(1, y) \in I(a) \times I(b)$ is not in the acceptability degree space (but still lies inside the blue area).*
- *Finally, if $I(b) = [x, y]$ s.t. $U < x \leq y$, then the ICAF is irrational because every point in $I(a) \times I(b)$ lies in the pink area (and thus above the blue curve).*

Note that intervals on acceptability degree cannot be directly converted into intervals for initial weights. This is due to the inter-dependencies between arguments in the final acceptability degree computation process. For example, if $I(a) = [0.8, 1]$ and $I(b) = [0, 0.5]$, one would think that the initial weights of a lies between 0.8 and 1 whereas the one of b would lie within 0 and 1. However, if $w(a) = 0.8$ and $w(b) = 1$, we would have $\sigma^{\texttt{Hbs}}_{\mathbf{wAF}}(b) \notin I(b)$.

If $(\mathbf{ICAF}, \sigma)$ is fully rational, then there is a weighting for any final acceptability degrees within the intervals and is thus *achievable*. In contrast, if $(\mathbf{ICAF}, \sigma)$ is rational but not fully rational, then some combination of final acceptability degrees within the interval will not be achievable. ϵ-rationality serves as a measure for "*how close to fully rational*" $(\mathbf{ICAF}, \sigma)$ is.

Definition 9. *Given $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$, a weighted gradual semantics σ and $0 \leq \epsilon$, we say $(\mathbf{ICAF}, \sigma)$ is ϵ-rational if for all $a \in \mathcal{A}$, it holds that $((\mathcal{A}, \mathcal{R}, I_a), \sigma)$ is rational, where for all $a' \in \mathcal{A}, a' \neq a$, $I_a(a') = I(a')$ and $I_a(a) = [\max(\min(I(a)), max(I(a)) - \epsilon), max(I(a))]$.*

The condition considers a set of induced ICAFs (one per argument) where, for each ICAF, a specific argument's interval is modified so that its lower bound is equal to its upper bound minus ϵ (and the remaining arguments are unchanged). If all such induced ICAFs are rational, then — in combination with the first condition — the original ICAF is ϵ-rational. Intuitively, this means that the shape $\times_{a \in \mathcal{A}} I(a)$ is not "too far" (i.e., within ϵ) from the acceptability degree space on all axes.

Observe that a fully rational ICAF must be 0-rational, any ϵ-rational (where $\epsilon > 0$) ICAF which is not ϵ'-rational (for any $\epsilon' < \epsilon$) is rational (but not fully rational), while an irrational ICAF is not ϵ-rational for any ϵ. Note that these are not "if and only if" relationships. Thus, a 0-rational ICAF is not necessarily fully rational, but is always rational.

Proposition 1. *If $(\mathbf{ICAF}, \sigma)$ is ϵ-rational, then it is rational.*

Proof. Let $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$ and assume $(\mathbf{ICAF}, \sigma)$ is ϵ-rational. For any $a \in \mathcal{A}$, if (1) $0 \leq \epsilon \leq \max(I(a)) - \min(I(a))$, we have that $((\mathcal{A}, \mathcal{R}, I_a), \sigma)$ is rational. Thus, there exists $w : \mathcal{A} \to [0, 1]$ s.t. $\sigma_{(\mathcal{A},\mathcal{R},w)}(a) \in [\max(I(a)) - \epsilon, \max(I(a))]$ and for all $b \in \mathcal{A} \setminus \{a\}, \sigma_{(\mathcal{A},\mathcal{R},w)}(b) \in I(b)$. Moreover, $\min(I(a)) \leq \max(I(a)) - \epsilon \leq \sigma_{(\mathcal{A},\mathcal{R},w)}(a) \leq \max(I(a))$, thus $\sigma_{(\mathcal{A},\mathcal{R},w)}(a) \in I(a)$. Thus, $(\mathbf{ICAF}, \sigma)$ is rational. If (2) $\epsilon > \max(I(a)) - \min(I(a))$, then $((\mathcal{A}, \mathcal{R}, I_a), \sigma) = ((\mathcal{A}, \mathcal{R}, I), \sigma)$ is rational. □

Rationality captures whether some part of the interval is in the acceptability degree space. A stronger concept is one of *refinement*. When refining intervals, we seek to determine whether the intervals can be tightened without losing any part of the acceptability degree space.

Definition 10. *Consider $(\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I), \sigma)$, we say that $\mathbf{ICAF}' = (\mathcal{A}, \mathcal{R}, I')$ is a* refinement *of $\mathbf{ICAF}$ (w.r.t. σ) iff (1) for all $a \in \mathcal{A}$, $I'(a) \subseteq I(a)$; and (2) for all $w : \mathcal{A} \to [0, 1]$ s.t. for all $a \in \mathcal{A}$, $\sigma_{(\mathcal{A},\mathcal{R},w)}(a) \in I(a)$, then $\sigma_{(\mathcal{A},\mathcal{R},w)}(a) \in I'(a)$. A refinement $\mathbf{ICAF}' = (\mathcal{A}, \mathcal{R}, I')$ of $\mathbf{ICAF}$ is called non-trivial if there exists some argument $a \in \mathcal{A}$ s.t. $I'(a) \subset I(a)$.*

Condition (2) is important and specifies that a refinement only removes unachievable acceptability degrees. A refinement thus shrinks intervals while ensuring that any achievable acceptability degrees remain.

Proposition 2. *If $(\mathbf{ICAF}, \sigma)$ is rational and $\mathbf{ICAF}'$ is a refinement of it, then $(\mathbf{ICAF}', \sigma)$ is rational.*

Proof. Assume that $(\mathbf{ICAF}, \sigma)$ is rational, then there is a weighting $w : \mathcal{A} \to [0, 1]$ such that for all $a \in \mathcal{A}, \sigma_{(\mathcal{A},\mathcal{R},w)}(a) \in I(a)$. From the second item of Definition 10, we get that for all $a \in \mathcal{A}, \sigma_{(\mathcal{A},\mathcal{R},w)}(a) \in I'(a)$. This means that $(\mathbf{ICAF}', \sigma)$ is rational. □

We now show an important intermediate result which informs us on the shape of the "acceptability degree space" of argumentation graphs, via the notion of *axial-radiality* for a weighted gradual semantics. Intuitively, an axial-radial weighted gradual semantics remains within the acceptability degree space if moving from a point within the space towards the origin. Satisfying axial-radiality means that a refinement can only reduce the upper bound of an interval.

Definition 11. *A weighted gradual semantics σ obeys* axial-radiality *iff for any $\mathbf{wAF} = (\mathcal{A}, \mathcal{R}, w)$, corresponding $\sigma_{\mathbf{wAF}}$ and $a \in \mathcal{A}$ then $\forall \varepsilon \in [0, \sigma_{\mathbf{wAF}}(a)]$, $((\mathcal{A}, \mathcal{R}, I), \sigma)$ is rational, where $I(a) = [\sigma_{\mathbf{wAF}}(a) - \varepsilon, \sigma_{\mathbf{wAF}}(a) - \varepsilon]$ and $\forall b \in \mathcal{A} \setminus \{a\}, I(b) = [\sigma_{\mathbf{wAF}}(b), \sigma_{\mathbf{wAF}}(b)]$.*

For weighted gradual semantics satisfying axial-radiality, the corner points of the hyper-rectangle induced by the intervals over arguments can be used to determine whether the pair ICAF/semantics is or is not (fully) rational.

Proposition 3. *Let $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$, where $\mathcal{A} = \{a_1, \ldots, a_n\}$, and let σ be semantics satisfying axial-radiality. For each i, define $Z_i = \{\min(I(a_i)), \max(I(a_i))\}$ and $Z = Z_1 \times Z_2 \times \cdots \times Z_n \subseteq [0, 1]^n$ be the set of all corner points of the hyper-rectangle induced by I. It holds that (1) if $Z \subseteq D_\sigma^{(\mathcal{A},\mathcal{R})}$, then $(\mathbf{ICAF}, \sigma)$ is fully rational; (2) else if $Z \cap D_\sigma^{(\mathcal{A},\mathcal{R})} = \emptyset$, then $(\mathbf{ICAF}, \sigma)$ is irrational; and (3) otherwise, $(\mathbf{ICAF}, \sigma)$ is rational but not fully rational.*

Proof. The point $(\max(I(a_1)), \ldots, \max(I(a_n))) \in Z$ is in the acceptability degree space and from the definition of axial-radiality, we conclude that every point $(d_1, d_2, \ldots, d_n)$, where $d_i \in I(a_i)$, is in the acceptability degree space, thus, $(\mathbf{ICAF}, \sigma)$ is fully rational.

Assume that none of the points of Z are in the acceptability degree space but that $(\mathbf{ICAF}, \sigma)$ is rational. This means that there exists $P = (d_1, d_2, \ldots, d_n)$, where $d_i \in I(a_i)$,

such that P is in the acceptability degree space. Using axial-radiality, we conclude that $(\min(a_1), \ldots, \min(a_n))$ is also in the acceptability degree space, contradiction.

Since one, but not all points, inside the acceptability space, $(\mathbf{ICAF}, \sigma)$ is rational but not fully rational. □

Item 3 of Proposition 3 does not mean that there exists a non-trivial refinement. Intuitively, a refinement exists if (at least) an entire face of an ICAF lies outside the acceptability degree space and at least one corner point lies within it.

Proposition 4. *Let $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$, $\mathcal{A} = \{a_1, \ldots, a_n\}$, a weighted gradual semantics σ satisfying axial-radiality, and Z be the set of all corner points of the hyper-rectangle induced by I s.t. $(\mathbf{ICAF}, \sigma)$ is rational. There is a non trivial refinement of $\mathbf{ICAF}$ (w.r.t. σ) iff there exists a set $X_i \subseteq Z$ composed of 2^{n-1} corner points s.t. every $(x_1, \ldots, x_n) \in X_i$ is not in the acceptability degree space and $x_i = \max(I(a_i))$.*

Proof. The sketch of the proof is as follows. $\Leftarrow$ Let us assume that X_i exists such that none of the points in X_i are in the acceptability degree space. Let P be the bounded hyperplane defined by the points in X_i. We know that none of the points in P are in the acceptability degree space because of axial-radiality (otherwise, one of the points in X_i would be in the acceptability degree space). There exists $\varepsilon > 0$ such that $X_i' = \{(x_1, \ldots, x_{i-1}, x_i - \varepsilon, x_{i+1}, \ldots, x_n) \mid (x_1, \ldots, x_n) \in X_i\}$ and all points in X_i' are also not in the acceptability degree space. It is easy to see that $\mathbf{ICAF}' = (\mathcal{A}, \mathcal{R}, I')$, where $I'(a) = [\min(a), x_i - \varepsilon]$ if $a = a_i$ and $I'(a) = I(a)$ otherwise, is a non trivial refinement of $\mathbf{ICAF}$.

$\Rightarrow$ Assume that $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$ has a non trivial refinement $\mathbf{ICAF}' = (\mathcal{A}, \mathcal{R}, I')$ such that $\forall a \in \mathcal{A}, I'(a) \subseteq I(a)$. Let a_i be such that $I'(a_i) \subset I(a_i)$ (this must be the case as the refinement is non-trivial). Let $X_i = \{(x_1, x_2, \ldots, x_{i-1}, \max(I(a_i)), \ldots, x_n) \mid \text{where for any } j \neq i,\ x_j \in \{\max(I(a_j)), \min(I(a_j))\}\}$ be the face of the hyper-rectangle in direction i, none of the points of X_i can be in the acceptability degree space, otherwise $\mathbf{ICAF}' = (\mathcal{A}, \mathcal{R}, I')$ would not be a refinement. □

Given a rational $(\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I), \sigma)$ and two of its refinements $\mathbf{ICAF}_1$ and $\mathbf{ICAF}_2$, $\mathbf{ICAF}_1$ is a *better refinement* than $\mathbf{ICAF}_2$ iff $\mathbf{ICAF}_1$ is a refinement of $\mathbf{ICAF}_2$.

3.1 Determining the Rationality of ICAFs

In this section, we focus on several popular weighted gradual semantics defined in the literature and identify properties of ICAFs with regards to these weighted gradual semantics.

Proposition 5. *Let $\mathbf{ICAF}$ be an arbitrary ICAF, it holds that (1) if the pair $(\mathbf{ICAF}, \sigma^{\texttt{Car}})$ is a rational then $(\mathbf{ICAF}, \sigma^{\texttt{Hbs}})$ is rational; and (2) if the pair $(\mathbf{ICAF}, \sigma^{\texttt{Hbs}})$ is a rational then $(\mathbf{ICAF}, \sigma^{\texttt{Max}})$ is rational.*

Proof. Let $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$ such that $(\mathbf{ICAF}, \sigma^{\texttt{Hbs}})$ is rational. There exists $w : \mathcal{A} \to [0, 1]$ such that $\forall a \in \mathcal{A}$, $\sigma^{\texttt{Hbs}}_{(\mathcal{A},\mathcal{R},w)}(a) \in I(a)$. [Oren *et al.*, 2022a] showed that $D^{\mathbf{AF}}_{\sigma^{\texttt{Max}}} \subseteq D^{\mathbf{AF}}_{\sigma^{\texttt{Hbs}}} \subseteq D^{\mathbf{AF}}_{\sigma^{\texttt{Car}}}$. Thus, there is $w' : \mathcal{A} \to [0, 1]$ such that $\forall a \in \mathcal{A}, \sigma^{\texttt{Max}}_{(\mathcal{A},\mathcal{R},w')}(a) = \sigma^{\texttt{Hbs}}_{(\mathcal{A},\mathcal{R},w)}(a)$ (and thus $\sigma^{\texttt{Max}}_{(\mathcal{A},\mathcal{R},w')}(a) \in I(a)$). This generalises to the relation between weighted card-based and weighted h-categorizer. □

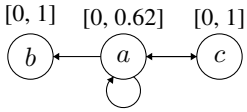


Figure 2: ICAF with 3 arguments.

Note that the inverse of Proposition 5 is not true.

Proposition 6. *Let $\sigma \in \{\sigma^{\texttt{Hbs}}, \sigma^{\texttt{Max}}, \sigma^{\texttt{Car}}\}$. σ satisfies axial-radiality.*

Proof. From [Oren *et al.*, 2022a], we know that – given a set of arguments and attacks between them – there is a one-to-one equivalence between a vector of weights ($\overrightarrow{w}$) and a vector of acceptability degrees ($\overrightarrow{X}$). For the weighted h-categorizer, one can go from the latter to the former using the equation $\overrightarrow{w} = \overrightarrow{X} + \mathbb{M}\mathbb{A}\overrightarrow{X}$, where $\mathbb{M}$ is the diagonal matrix with the acceptability degrees and $\mathbb{A}$ is the adjacency matrix. From this equation, one can check whether a vector of acceptability degrees is achievable or not by computing the vector of weights and checking if its values are between 0 and 1.

Consider an argumentation framework $\mathbf{AF} = (\mathcal{A}, \mathcal{R}, w)$ and corresponding $\sigma^{\texttt{Hbs}}$, vectors $\overrightarrow{w}$, $\overrightarrow{X}$, and matrices $\mathbb{M}$, $\mathbb{A}$.

Now, take any $a \in \mathcal{A}$ and $\varepsilon \in [0, \sigma_{\mathbf{AF}}(a)]$ and compute the vector $\overrightarrow{X'}$ where the value at the index corresponding to a is updated to $\sigma_{\mathbf{AF}}(a) - \varepsilon$. Similarly, we obtain $\mathbb{M}'$. Observe that $\overrightarrow{w'} = \overrightarrow{X'} + \mathbb{M}'\mathbb{A}\overrightarrow{X'}$ must be positive. Moreover, $\overrightarrow{w'} \leq \overrightarrow{w}$, meaning that $\overrightarrow{w'}$ has values ≤ 1.Thus, $((\mathcal{A}, \mathcal{R}, I), \sigma^{\texttt{Hbs}})$ is rational. From Prop. 5, we conclude the same for $\sigma^{\texttt{Max}}$. □

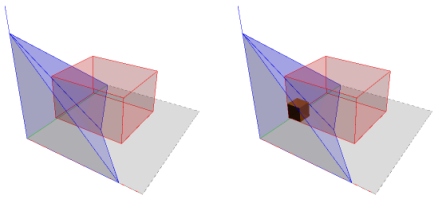
The next proposition answers the question *how can we determine if a pair $(\mathbf{ICAF}, \sigma)$ is rational/fully rational?*

Proposition 7. *It holds that $((\mathcal{A}, \mathcal{R}, I), \sigma)$ is rational iff $((\mathcal{A}, \mathcal{R}, I'), \sigma)$ is rational, where $\sigma \in \{\sigma^{\texttt{Hbs}}, \sigma^{\texttt{Max}}, \sigma^{\texttt{Car}}\}$ and for all $a \in \mathcal{A}$, $I'(a) = [\min(I(a)), \min(I(a))]$.*

Proof. We split this proof in two parts. ($\Leftarrow$) Assume that $((\mathcal{A}, \mathcal{R}, I'), \sigma)$ is rational, then $((\mathcal{A}, \mathcal{R}, I), \sigma)$ is also rational (by definition). ($\Rightarrow$) We show that if $((\mathcal{A}, \mathcal{R}, I), \sigma)$ is rational then $((\mathcal{A}, \mathcal{R}, I'), \sigma)$ is rational. Assume that $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$ and w such that $\mathbf{AF} = (\mathcal{A}, \mathcal{R}, w)$ and $\sigma_{\mathbf{AF}}(a) \in I(a)$, $\forall a \in \mathcal{A}$. By Proposition 6, $((\mathcal{A}, \mathcal{R}, I_1), \sigma)$ is rational, where $I_1(a_1) = [\min(I(a_1)), \min(I(a_1))]$ and $\forall b \in \mathcal{A} \setminus \{a_1\}$, $I_1(b) = I(b)$. Repeating this $n - 1$ more times proves that $((\mathcal{A}, \mathcal{R}, I'), \sigma)$ is rational. □

Proposition 7 yields a polynomial time procedure to determine whether an ICAF is rational (for a semantics) by checking whether the minimum values of the final acceptability degrees in its interval give an appropriate initial weight vector.

Corollary 8. *Let $((\mathcal{A}, \mathcal{R}, I))$ be an ICAF s.t. for all $a \in \mathcal{A}, \min(I(a)) = 0$. Then, for $\sigma \in \{\sigma^{\texttt{Hbs}}, \sigma^{\texttt{Max}}, \sigma^{\texttt{Car}}\}$, $((\mathcal{A}, \mathcal{R}, I), \sigma)$ is rational.*



(a) Proposition 7 (b) ε-refinement

Figure 3: Illustration of concepts.

Figure 3a illustrates the intuition of Proposition 7. Each axis represents the values of the degree of one argument (the figure represents a scenario with 3 arguments). The blue shape is the acceptability degree space, and the red shape corresponds to the intervals for each argument given by I. $((\mathcal{A}, \mathcal{R}, I), \sigma)$ is rational iff there is an intersection between the red shape and the blue shape. This happens iff the "minimal" corner of the red shape is inside the blue shape.

Proposition 9. *$((\mathcal{A}, \mathcal{R}, I), \sigma)$ is fully rational iff $((\mathcal{A}, \mathcal{R}, I'), \sigma)$ is rational, where $\sigma \in \{\sigma^{\texttt{Hbs}}, \sigma^{\texttt{Max}}, \sigma^{\texttt{Car}}\}$ and for all $a \in \mathcal{A}$, $I'(a) = [\max(I(a)), \max(I(a))]$.*

Proof. The proof is similar to the one of Proposition 7. □

3.2 Refining Intervals

We now consider the problem of obtaining the "best" refinement of a rational interval constrained argumentation framework w.r.t. some weighted gradual semantics. Such a refinement is an ϵ-rational refinement for which no elements of the acceptability degree space are lost, and which cannot be further refined without losing some elements of the acceptability degree space. Algorithm 1 describes how such a refinement can be identified. Lines 1-3 check that the condition of Proposition 4 is satisfied and that we can find a non-trivial refinement. Lines 4-13 modify the argument intervals using the bisection method until we obtain an ε-refinement. Note that the output is not fully rational due to line 8. The use of the bisection method means that the distance between l and r is halved at each iteration, leading to convergence in approximately $\log \varepsilon$ iterations. Figure 3b illustrates the best refinement (the small orange cuboid) of an ICAF (red cuboid).

Proposition 10. *Given $(\mathbf{ICAF}, \sigma)$ (s.t. σ is axial-radial) and $\varepsilon > 0$ as input to Algorithm 1, it outputs an 2ε-rational refinement $\mathbf{ICAF}'$ of $(\mathbf{ICAF}, \sigma)$.*

Proof. Given as input a rational pair $(\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I), \sigma)$, the algorithm modifies I into I_{out}. If $\mathbf{ICAF}$ does not have a non-trivial refinement (according to Proposition 4), we get $I_{out} = I$, showing that the initial ICAF is already the best refinement we can obtain (and trivially a 2ε-rational refinement of itself).

If instead a non-trivial refinement exists, we need to show that the two conditions of Definition 10 hold. Notice that $\forall a \in \mathcal{A}$, $I_{out}(a) \subseteq I(a)$ (satisfying item (1) of Definition 10). At the end of the algorithm, there exists a non-empty $X \subseteq \mathcal{A}$ s.t. $\forall x \in X$, $\mathbf{ICAF}_x = ((\mathcal{A}, \mathcal{R}, I''), \sigma)$ is irrational with $I''(x) = [m_x, m_x]$ (line 8) and $\forall a \in \mathcal{A} \setminus X$, $I_{out}(a) = I(a)$. Since $\forall x \in X$, $I_{out}(x) = [\min(I(x)), m_x]$, and σ satisfies axial-radiality, item (2) of Definition 10 is satisfied (no achievable acceptability degrees were removed). Thus, $(\mathcal{A}, \mathcal{R}, I_{out})$ is a refinement of $(\mathcal{A}, \mathcal{R}, I)$ (w.r.t. σ).

The refinement is 2ε-rational. The *found* variable in the while loop is only set to *True* when the condition on line 11 is satisfied. Thus, the interval is at most 2ε in size, ensuring that $(\mathcal{A}, \mathcal{R}, I_{out})$ is 2ε-rational.

□

Algorithm 1 Computing the best refinement

Require: A rational pair $((\mathcal{A}, \mathcal{R}, I), \sigma)$ and $\varepsilon > 0$.

1: **for** $a \in \mathcal{A}$ **do**
2: $\quad \mathbf{ICAF}_a \leftarrow ((\mathcal{A}, \mathcal{R}, I'), \sigma)$ where for all $b \in \mathcal{A} \setminus \{a\}, I'(b) = I(b)$ and $I'(a) = [\max(I(a)), \max(I(a))]$
3: $\quad$ **if** $\mathbf{ICAF}_a$ is irrational **then**
4: $\quad\quad l \leftarrow \min(I(a))$, $r \leftarrow \max(I(a))$, $found \leftarrow \bot$
5: $\quad\quad$ **while** $\neg found$ **do**
6: $\quad\quad\quad m \leftarrow (l + r)/2$
7: $\quad\quad\quad \mathbf{ICAF}'_a \leftarrow ((\mathcal{A}, \mathcal{R}, I''), \sigma)$ where for all $b \in \mathcal{A} \setminus \{a\}, I''(b) = I(b)$ and $I''(a) = [m, m]$
8: $\quad\quad\quad$ **if** $\mathbf{ICAF}'_a$ is irrational **then**
9: $\quad\quad\quad\quad r \leftarrow m$
10: $\quad\quad\quad\quad$ **if** $(r - l)/2 < \varepsilon$ **then**
11: $\quad\quad\quad\quad\quad I(a) \leftarrow [\min(I(a)), m]$
12: $\quad\quad\quad\quad\quad found \leftarrow \top$
13: $\quad\quad\quad\quad$ **else** $l \leftarrow m$

3.3 Correcting Irrationality

Consider the situation where an ICAF under a given weighted gradual semantics is irrational. We may determine some minimal interval changes required to achieve rationality.

Strategy 1. We associate a cost with modifying an argument, according to the cost function $cost : \mathcal{A} \to \mathbb{N}^*$. To change an argument a by Δ therefore costs $cost(a) * \Delta$. We can thus consider a line intersecting the point $S = (\min(I(a_1)), \ldots, \min(I(a_n)))$ with gradient vector $1/cost(a_i)$ for component i. Moving along this line has an equivalent cost for all arguments. Thus, our strategy consists of starting at point S and creating a new ICAF (with interval function I') whose minimum interval moves along the negation of the gradient until this new ICAF is rational, or until $\min(I'(a_i))$ is equal to 0 (for some i). In the former case, we return the new ICAF, while in the latter, we continue moving down while keeping $\min(I'(a_i)) = 0$. This approach is described in Algorithm 2. It uses the bisection method to efficiently traverse the gradient line. Lines 1-4 compute the start point of the line (the end point lies at 1). In lines 5–8, we use the bisection method to identify the extreme point where the resultant intervals are rational, taking the possibility of $\min(I'(a))$ being below 0 into account (line 7). This strategy clearly has polynomial complexity.

Strategy 2. This strategy exhaustively applies Strategy 1 but restricts it so that only the intervals of a subset of arguments $X \subseteq \mathcal{A}$ can be modified. When all argument subsets have been explored, the cost-minimising subset is selected.

Algorithm 2 Computing rational ICAF using strategy 1

Require: An irrational $((\mathcal{A}, \mathcal{R}, I), \sigma)$, s.t. σ satisfies axiality and radiality and $\varepsilon > 0$.

$l \leftarrow 1, u \leftarrow 1$
for all $a \in \mathcal{A}$ **do**
  $t \leftarrow 1 - (1/costs(a)) * \min(I(a))$
  if $t \leq l_0$ **then** $l \leftarrow t$
while $u - l \geq \varepsilon$ **do**
  $m \leftarrow (l + u)/2$
  $\mathbf{ICAF}' \leftarrow ((\mathcal{A}, \mathcal{R}, I'), \sigma)$ s.t. $\forall a \in \mathcal{A}$, $I'(a) = [\max(0, \frac{m}{costs(a)} + min(I(a)) - costs(a)), \max(I(a))]$
  if $\mathbf{ICAF}'$ is rational **then** $l \leftarrow m$ **else** $u \leftarrow m$
return $((\mathcal{A}, \mathcal{R}, I''), \sigma)$ where for every $a \in \mathcal{A}, I''(a) = [\max(0, l/costs(a) + \min(I(a))), \max(I(a))]$

Other strategies. We can consider two special cases of the previous strategies that do not require argument costs. First, we may consider the case where $costs(a_i) = 1$, for all $1 \leq i \leq n$. Second, we may move along the line segment between the origin and S, in which case our gradient vector has $\min(I(a_i))$ as its i-th component. This latter approach has the benefit of no argument having 0 as its minimum element in the interval (unless all arguments have 0 as the minimum interval element). Such a strategy can be captured using a cost function, and therefore, we do not consider these further.

The aim of all strategies is to identify the closest point to S in the acceptability degree space. Cost-aware strategies change the metric by which this distance is measured. The need for strategies arises as — in general — one cannot identify the closest rational point to another point analytically.

4 Evaluation

To evaluate the strategies from Section 3.3, we generated 40 Erdös-Rényi random graphs[2] for each number of arguments between 4 and 14, with a 0.5 likelihood for edge creation. Each such graph yielded a single ICAF, and for each such ICAF, we initialized I s.t., $I(a) = [x, 1]$, where x is picked randomly from $[0.8, 1]$, ensuring that $((\mathcal{A}, \mathcal{R}, I), \sigma)$ is likely irrational according to the chosen weighted gradual semantics. Argument costs were drawn from $[1, 11]$.

Figures 4 (a-c) illustrate the cost for our cost-aware strategies using the weighted h-categoriser, card-based and max-based semantics. For the weighted h-categoriser semantics, Strategy 2 outperforms Strategy 1, while the advantage of the former strategy shrinks for the weighted card-based and max-based semantics as the number of arguments grows. Figures 4 (d-f) plot the total number of arguments modified to achieve minimal cost via Strategy 2. In all cases, the number of arguments grows linearly. However, we can see that the variance in the number of arguments modified remains large for the weighted h-categoriser, while shrinking for weighted max-based semantics. For the weighted card-based semantics, the variance begins small and quickly approaches 0 as the number of arguments increases. We hypothesize that the explicit consideration of the number of attacks in weighted-card based semantics reduces the impact of arguments' initial weight on rationality, while the use of only maximally acceptable arguments in the weighted max-based semantics causes a similar effect. The exponential complexity of Strategy 2 makes it unsuitable for large ICAFs. Our results suggest that for the weighted card-based and max-based semantics, the impact of using Strategy 1 is small, and that for the weighted h-categoriser semantics, Strategy 1 performs well.

[2]Note that similar results were obtained for other networks with no islands; space constraints prevent us from including other results.

5 Sampling Initial Weights and Insights

Using our pipeline, users can specify intervals of acceptability degrees as ICAFs, and refine or correct the elicited intervals for each argument. Now, to obtain a possible value for the initial weights associated with each argument, one can use a sampling approach to efficiently get a weighting function w.

Definition 12. *Let $(\mathbf{ICAF}, \sigma)$ be rational, a weighting function w can be sampled (from $\mathbf{ICAF}$ w.r.t. σ) iff for all $a \in \mathcal{A}, \sigma_{(\mathcal{A},\mathcal{R},w)}(a) \in I(a)$. The set of all weighting functions that can be sampled is denoted $S(\mathbf{ICAF}, \sigma)$. We denote by $W_{\mathcal{A}}$, the set of all weighting functions from $\mathcal{A}$.*

We can determine the probability of specific preferences in initial weights between arguments using ranking queries.

Definition 13. *Given a set of arguments $\mathcal{A}$, a ranking query (on $\mathcal{A}$) is a Boolean combination of expressions of the form $a \leq b$ with $a, b \in \mathcal{A}$. We use Greek letters $\alpha, \beta, \gamma, \ldots$ to denote ranking queries.*

For a weighting function $w : \mathcal{A} \to [0, 1]$, we say that $w \models a \leq b$ iff $w(a) \leq w(b)$. Similarly, $w \models \alpha \vee \beta$ iff $w \models \alpha$ or $w \models \beta$. We can now define how probabilities on ranking queries are calculated.

Definition 14. *Consider $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$, a weighted gradual semantics σ s.t. $(\mathbf{ICAF}, \sigma)$ is rational, and a ranking query α. Let $Y = S(\mathbf{ICAF}, \sigma)$, the probability of α is $P_{\mathbf{ICAF},\sigma}(\alpha) = \frac{|\{w \in Y | w \models \alpha\}|}{|Y|}$ if $Y \neq \emptyset$ and 0 otherwise.*

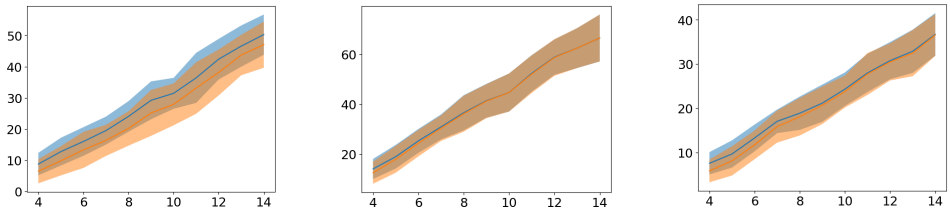
In Definition 14, the numerator is inferior or equal to the denominator, and thus $P_{\mathbf{ICAF},\sigma}(\alpha) \in [0, 1]$. By sampling, we can approximate $P_{\mathbf{ICAF},\sigma}(\alpha)$.

Example 2. *Let us consider $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$ as displayed in Figure 2. We get that $P_{\mathbf{ICAF},\sigma^{\text{Hbs}}}(a \leq b) \approx 0.608$, $P_{\mathbf{ICAF},\sigma^{\text{Hbs}}}(a \leq b \wedge b \leq c) \approx 0.208$, $P_{\mathbf{ICAF},\sigma^{\text{Hbs}}}(c \leq a \wedge a \leq b) \approx 0.167$, and $P_{\mathbf{ICAF},\sigma^{\text{Hbs}}}(a \leq c \wedge c \leq b) \approx 0.233$. Note that $P_{\mathbf{ICAF},\sigma^{\text{Hbs}}}(a \leq b) = P_{\mathbf{ICAF},\sigma^{\text{Hbs}}}(a \leq b \wedge b \leq c) + P_{\mathbf{ICAF},\sigma^{\text{Hbs}}}(c \leq a \wedge a \leq b) + P_{\mathbf{ICAF},\sigma^{\text{Hbs}}}(a \leq c \wedge c \leq b)$.*

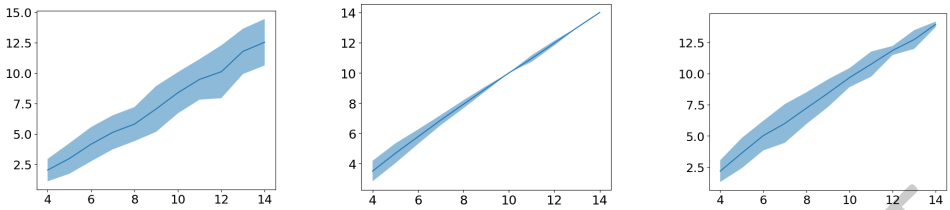
It is also possible to get insights on the ICAF by computing the fraction of the initial weight space (resp. acceptability degree space) that satisfies the ICAF constraints.

Definition 15. *Consider $\mathbf{ICAF} = (\mathcal{A}, \mathcal{R}, I)$ and a weighted gradual semantics σ, with $\mathcal{A} = \{a_1, \ldots, a_n\}$. The fraction of the initial weight space (resp. degree space) that satisfies the ICAF constraints is $F^{init}_{\mathbf{ICAF},\sigma} = \frac{|S(\mathbf{ICAF},\sigma)|}{|W_{\mathcal{A}}|}$ (resp. $F^{deg}_{\mathbf{ICAF},\sigma} = \frac{|\{P=(p_1,\ldots,p_n) \in D^{\mathbf{AF}}_{\sigma} | \forall 1 \leq i \leq n, p_i \in I(a_i)\}|}{|D^{\mathbf{AF}}_{\sigma}|}$).*

Example 3 highlights the difference between these two fractions as $F^{init}_{\mathbf{ICAF},\sigma}$ samples the initial weight space while $F^{deg}_{\mathbf{ICAF},\sigma}$ samples the degree space.



(a) W. h-categoriser (b) W. card-based (c) W. max-based



(d) W. h-categoriser (e) W. card-based (f) W. max-based

Figure 4: (a–c) Cost (y-axis) achieved by Strategy 1 (blue) and 2 (orange) w.r.t. number of arguments (x-axis); shaded regions denote ± 1 standard deviation. (d–f) Number of arguments used (y-axis) by Strategy 2 w.r.t. the total number of arguments, averaged over 40 runs.

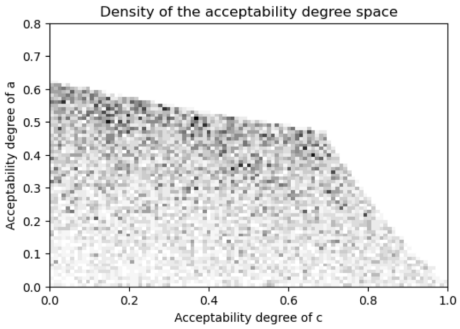


Figure 5: Density of the acceptability degree space for values of b between 0.30 and 0.31. Darker shading indicates higher density.

Example 3. *Consider* $\mathbf{ICAF} = (\{a, b, c\}, \mathcal{R}, I')$*, where* $I'(a) = [0.2, 0.4]$ *and* $I'(b) = I'(c) = [0, 1]$*. Using* $1\,000$ *sampled points and repeating the experiment* $5\,000$ *times, the mean values of* $F^{init}_{\mathbf{ICAF},\sigma}$ *and* $F^{deg}_{\mathbf{ICAF},\sigma}$ *are 0.38 and* 0.34*, respectively. The difference between these means is significant* ($p \approx 0$) *using an independent two-sample t-test).*

This is because, in general, the acceptability degree space does not have a "homogeneous density". Figure 5 demonstrates this by uniformly sampling 10^6 *initial weights and plotting the number of points where the degree of* b *is between* 0.3 *and* 0.31 *(yielding a slice) for the sampled weights.*

Sometimes, $F^{init}_{\mathbf{ICAF},\sigma} = F^{deg}_{\mathbf{ICAF},\sigma}$. Two such cases are in ICAFs without attacks, and when the acceptability degree space is fully included in the constraints (i.e., $\forall P = (p_1, \ldots, p_n) \in D^{\mathbf{AF}}_{\sigma}$, we have $p_i \in I(a_i)$, for every $1 \leq i \leq n$). In the latter case, we have $F^{init}_{\mathbf{ICAF},\sigma} = F^{deg}_{\mathbf{ICAF},\sigma} = 1$.

6 Related Work

In [Espinoza *et al.*, 2023], the authors introduce credal support argumentation frameworks. Here each argument is associated with a credal set and an imprecise base score obtained from its credal set. While they only consider a support relation, they show how to compute the imprecise strength of arguments (an interval).The epistemic approach to probabilistic argumentation [Hunter *et al.*, 2021] seeks valid probabilities for arguments given some properties, similar to our intervals. In fuzzy argumentation, legal argument weights can be computed according to the approach of [Wu *et al.*, 2016]. However, in all these cases, the properties of the interval are not explicitly considered, nor are correcting systems which do not comply with the underlying properties. Enforcement in abstract argumentation (e.g., [Baumann *et al.*, 2021]) considers how argumentation frameworks can be modified to guarantee an argument's status and refinement can be viewed similarly in a weighted gradual semantics context.

7 Conclusions and Future Work

In this paper, we use weighted gradual semantics to align user beliefs as intervals of acceptability degrees for arguments. From an initial set of acceptability degree intervals we introduced refinement (tightening interval bounds without losing valid solutions) and rationalisation (modifying bounds to include valid solutions not originally present). The latter was achieved via cost-aware heuristics.

We intend to pursue several avenues of future work. We plan to investigate how to sample the "best" valid initial weights, and present/explain them to a user, allowing them to identify those that they wish to use. In Definition 15, we provided two measures to reflect the rationality of an ICAF. In conjunction with other human studies, we will investigate which of these better aligns with human intuition. We also intend to further examine the theoretical properties of these measures. Finally, extending the formalism to capture intervals via probability distributions may also better capture human reasoning.

References

[Amgoud and Ben-Naim, 2013a] Leila Amgoud and Jonathan Ben-Naim. Ranking-Based Semantics for Argumentation Frameworks. In *Scalable Uncertainty Management - 7th International Conference, SUM 2013, Washington, DC, USA, September 16-18, 2013. Proceedings*, pages 134–147, 2013.

[Amgoud and Ben-Naim, 2013b] Leila Amgoud and Jonathan Ben-Naim. Ranking-based semantics for argumentation frameworks. In Weiru Liu, V. S. Subrahmanian, and Jef Wijsen, editors, *Scalable Uncertainty Management - 7th International Conference, SUM 2013, Washington, DC, USA, September 16-18, 2013. Proceedings*, volume 8078 of *Lecture Notes in Computer Science*, pages 134–147. Springer, 2013.

[Amgoud and Ben-Naim, 2018] Leila Amgoud and Jonathan Ben-Naim. Weighted Bipolar Argumentation Graphs: Axioms and Semantics. 2018.

[Amgoud *et al.*, 2022] Leila Amgoud, Dragan Doder, and Srdjan Vesic. Evaluation of argument strength in attack graphs: Foundations and semantics. *Artificial Intelligence*, 302:103607, January 2022.

[Baroni *et al.*, 2011] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. An introduction to argumentation semantics. *Knowledge Eng. Review*, 26(4):365–410, 2011.

[Baumann *et al.*, 2021] Ringo Baumann, Sylvie Doutre, Jean-Guy Mailly, and Johannes Peter Wallner. Enforcement in Formal Argumentation. *IfColog Journal of Logics and their Applications (FLAP)*, 8(6):1623–1678, July 2021. ISBN : 978-1-84890-371-5.

[Besnard and Hunter, 2001] Philippe Besnard and Anthony Hunter. A logic-based theory of deductive arguments. *Artif. Intell.*, 128(1-2):203–235, 2001.

[Bonzon *et al.*, 2018] Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, and Nicolas Maudet. Combining Extension-Based Semantics and Ranking-Based Semantics for Abstract Argumentation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Sixteenth International Conference, KR 2018, Tempe, Arizona, 30 October - 2 November 2018.*, pages 118–127, 2018.

[Bonzon *et al.*, 2023] Elise Bonzon, Jérôme Delobelle, Sébastien Konieczny, and Nicolas Maudet. An empirical and axiomatic comparison of ranking-based semantics for abstract argumentation. *J. Appl. Non Class. Logics*, 33(3-4):328–386, 2023.

[Dung, 1995] Phan Minh Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artif. Intell.*, 77(2):321–358, 1995.

[Espinoza *et al.*, 2023] Mariela Morveli Espinoza, Juan Carlos Nieves, and Cesar A. Tacla. A gradual semantics with imprecise probabilities for support argumentation frameworks. In Kai Sauerwald and Matthias Thimm, editors, *Proceedings of the 21st International Workshop on Non-Monotonic Reasoning co-located with the 20th International Conference on Principles of Knowledge Representation and Reasoning (KR 2023) and co-located with the 36th International Workshop on Description Logics (DL 2023), Rhodes, Greece, September 2-4, 2023*, volume 3464 of *CEUR Workshop Proceedings*, pages 84–93. CEUR-WS.org, 2023.

[Hunter *et al.*, 2021] Anthony Hunter, Sylwia Polberg, Nikolas Potyka, Tjitze Rienstra, and Mathhias Thimm. *Handbook of Formal Argumentation*, volume 2, chapter Probabilistic argumentation: A Survey. College Publications, 2021.

[IPCC, 2023] IPCC. Climate change 2023: Synthesis report, 2023.

[Oren *et al.*, 2022a] Nir Oren, Bruno Yun, Assaf Libman, and Murilo S. Baptista. Analytical solutions for the inverse problem within gradual semantics. *CoRR*, abs/2203.01201, 2022.

[Oren *et al.*, 2022b] Nir Oren, Bruno Yun, Srdjan Vesic, and Murilo S. Baptista. Inverse problems for gradual semantics. In Luc De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 2719–2725. ijcai.org, 2022.

[Rago *et al.*, 2016] Antonio Rago, Francesca Toni, Marco Aurisicchio, and Pietro Baroni. Discontinuity-Free Decision Support with Quantitative Argumentation Debates. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016.*, pages 63–73, 2016.

[Rahwan *et al.*, 2010] Iyad Rahwan, Mohammed Iqbal Madakkatel, Jean-François Bonnefon, Ruqiyabi Naz Awan, and Sherief Abdallah. Behavioral experiments for assessing the abstract argumentation semantics of reinstatement. *Cogn. Sci.*, 34(8):1483–1502, 2010.

[Vesic *et al.*, 2022] Srdjan Vesic, Bruno Yun, and Predrag Teovanovic. Graphical representation enhances human compliance with principles for graded argumentation semantics. In Piotr Faliszewski, Viviana Mascardi, Catherine Pelachaud, and Matthew E. Taylor, editors, *21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2022, Auckland, New Zealand, May 9-13, 2022*, pages 1319–1327. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 2022.

[Wu *et al.*, 2016] Jiachao Wu, Hengfei Li, Nir Oren, and Timothy J. Norman. Gödel fuzzy argumentation frameworks. In *COMMA*, volume 287 of *Frontiers in Artificial Intelligence and Applications*, pages 447–458. IOS Press, 2016.

[Yun *et al.*, 2020] Bruno Yun, Srdjan Vesic, and Madalina Croitoru. Ranking-based semantics for sets of attacking arguments. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference,*

IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020, pages 3033–3040. AAAI Press, 2020.