On Restricting the Impact of Self-Attacking Arguments in Gradual Semantics

Vivien Beuselinck, Jérôme Delobelle, Srdjan Vesic

On Restricting the Impact of Self-Attacking Arguments in Gradual Semantics

Vivien Beuselinck[1][0000−0002−5722−7025], Jérôme Delobelle[2][0000−0003−1691−4731], and Srdjan Vesic[3][0000−0002−4382−0928]

[1] Aniti, Université Fédérale, vivien@beuselinck.fr
[2] LIPADE, Université de Paris, France, jerome.delobelle@u-paris.fr
[3] CNRS, Univ. Artois, CRIL, France, vesic@cril.fr

Abstract. The issue of how a semantics should deal with self-attacking arguments was always a subject of debate amongst argumentation scholars. A consensus exists for extension-based semantics because those arguments are always rejected (as soon as the semantics in question respect conflict-freeness). In case of gradual semantics, the question is more complex, since other criteria are taken into account. A way to check the impact of these arguments is to use the principles (i.e. desirable properties to be satisfied by a semantics) from the literature. Principles like Self-Contradiction and Strong Self-Contradiction prescribe how to deal with self-attacking arguments. We show that they are incompatible with the well-known Equivalence principle (which is satisfied by almost all the existing gradual semantics), as well as with some other principles (e.g. Counting). This incompatibility was not studied until now and the class of semantics satisfying Self-Contradiction is under-explored. In the present paper, we explore that class of semantics. We show links and incompatibilities between several principles. We define a semantics that satisfies (Strong) Self-Contradiction and a maximal number of compatible principles. We introduce an iterative algorithm to calculate our semantics and prove that it always converges. We also provide a characterisation of our semantics. Finally, we experimentally show that our semantics is computationally efficient.

Keywords: Abstract argumentation · Gradual semantics · Self-attack.[4]

1 Introduction

Theory of computational argumentation allows to model exchange of arguments and conflicts between them. Although in most cases a conflict occurs between two arguments, sometimes an argument may conflict with itself. Such an argument is called a self-attacking argument. Discussion on how to deal with self-attacking arguments is often indirectly included in the problems of dealing with odd-length cycles, because a self-attack is the smallest odd-length cycle. However, in contrast to greater odd-length cycles, the presence of a self-attack is due to inconsistency in an argument itself.

[4] The final authenticated publication is available online at the following address : https://doi.org/10.1007/978-3-030-89391-0_8

In order to reason in presence of these arguments, several methods have been defined in abstract argumentation by proposing to deal with them directly [11,9,8,16] or indirectly [7]. These methods essentially concern extension-based semantics. In the context of ranking-based and gradual argumentation semantics [2,5], little research was conducted to find out how self-attacking arguments should be dealt with and what is the impact they have on the acceptability of other arguments. Existing studies are essentially done through the principle-based study of these semantics. Indeed, defining and studying principles drew attention of many scholars in this area.

Consider Equivalence, which is one of the well-known principles, stating that the acceptability degree of an argument should only depend on acceptability degrees of its direct attackers and consider the argumentation graph $\mathcal{F}_{ex}$ containing two arguments a and b, and where b is attacked by a self-attacking argument a (i.e., $\mathcal{F}_{ex} = (\{a, b\}, \{(a, a), (a, b)\})$). Equivalence implies that a and b should be equally acceptable because a and b are both attacked by a self-attacking argument. However, this is debatable, since the intuition behind a self-attacking argument is that it is inconsistent in one way or another so we would tend to accept b being attacked by a (which is self-attacking) rather than accepting a. Note that, under all semantics returning conflict-free extensions, a self-attacking argument is always rejected, i.e. it does not belong to any extension. Also, regarding the ranking-based and gradual semantics, it was pointed out that it would be natural to attach the worst possible rank to self-attacking arguments [19]. Furthermore, two principles were defined to formalise this intuition.

The first one is called Strong Self-Contradiction, and introduced by Matt and Toni [19]. It says that the acceptability degree of an argument must be 0 if and only if that argument is self-attacking. The second principle, called Self-Contradiction, was introduced by Bonzon et al. [12] and states that every self-attacking argument is strictly less acceptable than every non self-attacking argument. Consider the argumentation graph $\mathcal{F}_{ex}$ again and note that, under every semantics that satisfies Self-Contradiction, b is strictly more acceptable than a. This example shows that Equivalence and Self-Contradiction are not compatible, i.e. there exists no semantics that satisfies both of them.

To the best of our knowledge, there exists only one semantics (known as M&T) that satisfies Self-Contradiction and Strong Self-Contradiction. That semantics was introduced by Matt and Toni [19]. However, this semantics has a limitation that makes it inapplicable in practice. Namely, as noted by Matt and Toni themselves, as the space used to calculate the scores grows exponentially with the number of arguments, even with the optimisation techniques they used it did not scale to more than a dozen of arguments.

The research objective of the present paper is to study the under-explored family of semantics that satisfy Strong Self-Contradiction. Our goals are thus to identify which principles are (in)compatible with Strong Self-Contradiction and to define a semantics, which we call `nsa` (no self-attacks), that satisfies Strong Self-Contradiction as well as a maximal number of compatible principles. After introducing the formal setting and recalling the existing principles from the literature:

- We prove the incompatibilities between some of the principles, and identify a maximal set of principles that contains (Strong) Self-Contradiction;

- We introduce an iterative algorithm in order to define a new semantics and prove that it always converges. The acceptability of degree of each argument with respect to `nsa` is then defined as the limit of the corresponding sequence;
- We provide a characterisation of `nsa`, i.e. a declarative (non-iterative) definition and show that the two are equivalent: each semantics satisfying the declarative definition coincides with `nsa`;
- We check which principles are satisfied by `nsa` and compare it with the h-categorizer semantics [10] and the M&T semantics in terms of principle satisfaction;
- We formally prove that no semantics can satisfy a strict super-set of the set of principles satisfied by `nsa`;
- We experimentally show that `nsa` is computationally efficient and compare it with the M&T semantics and the h-categorizer semantics. The results confirm the hypothesis that the M&T semantics does not scale.

2 Formal Setting and Existing Semantics

An argumentation graph (AG) [17] is a directed graph $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ where $\mathcal{A}$ is a finite set of arguments and $\mathcal{R}$ a binary relation over $\mathcal{A}$, i.e. $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. For $a, b \in \mathcal{A}$, $(a, b) \in \mathcal{R}$ means that a attacks b. The notation $\mathtt{Att}_{\mathcal{F}}(a) = \{b \mid (b, a) \in \mathcal{R}\}$ represents the set of direct attackers of argument a. For two graphs $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ and $\mathcal{F}' = (\mathcal{A}', \mathcal{R}')$, we denote by $\mathcal{F} \otimes \mathcal{F}'$ the argumentation graph $\mathcal{F}'' = (\mathcal{A} \cup \mathcal{A}', \mathcal{R} \cup \mathcal{R}')$.
Dung's framework comes equipped with various types of semantics used to evaluate the arguments. These include the extension-based semantics (see [6] for an overview), the labelling-based semantics [14], the ranking-based semantics (see [12] for an overview) and the gradual semantics. We refer the reader to [13,1] for a complete overview of the existing families of semantics in abstract argumentation and the differences between these approaches (e.g., definition, outcome, application). In this article, we focus on gradual semantics which assign to each argument in an argumentation graph a score, called *acceptability degree*. This degree belongs to the interval $[0, 1]$. Higher degrees correspond to stronger arguments.

Definition 1 (Gradual semantics). *A gradual semantics is a function $\mathcal{S}$ which associates to any argumentation graph $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ a function $Deg^{\mathcal{S}}_{\mathcal{F}} : \mathcal{A} \to [0, 1]$. Thus, $Deg^{\mathcal{S}}_{\mathcal{F}}(x)$ represents the acceptability degree of $x \in \mathcal{A}$.*

In the rest of the section we recall two gradual semantics. We first introduce h-categorizer, which is one of the most studied gradual semantics and also satisfies a maximal compatible set of principles from the literature.[5] Then we introduce M&T semantics which is, to the best of our knowledge, the only semantics known in the literature to satisfy Self-Contradiction.

[5] formally: out of the principles from Section 3, no semantics satisfies a strict superset of the principles satisfied by h-categorizer.

2.1 h-categorizer Semantics

The h-categorizer semantics [10,20] uses a categorizer function to assign a value to each argument by taking into account the strength of its attackers, which itself takes into account the strength of its attackers, and so on.

Definition 2 (h-categorizer semantics). *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an argumentation graph. The h-categorizer semantics is a gradual semantics such that $\forall x \in \mathcal{A}$:*

$$Deg^{\mathtt{h}}_{\mathcal{F}}(x) = \frac{1}{1 + \sum_{y \in \mathtt{Att}_{\mathcal{F}}(x)} Deg^{\mathtt{h}}_{\mathcal{F}}(y)}$$

2.2 M&T Semantics

The gradual semantics introduced by Matt and Toni [19] computes the acceptability degree of an argument using a two-person zero-sum strategic game. For an AG $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ and an argument $x \in \mathcal{A}$, the set of strategies for the proponent is the set of all subsets of arguments that contain x: $S_P(x) = \{P \mid P \subseteq \mathcal{A}, x \in P\}$ and for the opponent it is the set of all subsets of arguments: $S_O = \{O \mid O \subseteq \mathcal{A}\}$. Given two strategies $X, Y \subseteq \mathcal{A}$, the set of attacks from X to Y is defined by $Y^{\leftarrow X}_{\mathcal{F}} = \{(x, y) \in X \times Y \mid (x, y) \in \mathcal{R}\}$. From this measurement, Matt and Toni define the notion of degree of acceptability of a set of arguments w.r.t. another one used to compute the reward of a proponent's strategy.

Definition 3 (Reward). *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an argumentation graph, $x \in \mathcal{A}$ be an argument, $P \in S_P(x)$ be a strategy chosen by the proponent and $O \in S_O$ be a strategy chosen by the opponent. The degree of acceptability of P w.r.t. O is $\phi(P, O) = \frac{1}{2}\left[1 + f(|O^{\leftarrow P}_{\mathcal{F}}|) - f(|P^{\leftarrow O}_{\mathcal{F}}|)\right]$ with $f(n) = \frac{n}{n+1}$. The reward of P over O, denoted by $r_{\mathcal{F}}(P, O)$, is defined by:*

$$r_{\mathcal{F}}(P, O) = \begin{cases} 0 & \text{iff } P \text{ is not conflict-free} \\ 1 & \text{iff } P \text{ is conflict-free and} \\ & \quad |P^{\leftarrow O}_{\mathcal{F}}| = 0 \\ \phi(P, O) & \text{otherwise} \end{cases}$$

Proponent and opponent have the possibility of using a strategy according to some probability distributions, respectively $p = (p_1, p_2, \ldots, p_m)$ and $q = (q_1, q_2, \ldots, q_n)$, with $m = |S_P|$ and $n = |S_O|$. For each argument $x \in \mathcal{A}$, the proponent's expected payoff $E(x, p, q)$ is then given by $E(x, p, q) = \sum_{j=1}^{n} \sum_{i=1}^{m} p_i q_j r_{i,j}$ with $r_{i,j} = r_{\mathcal{F}}(P_i, O_j)$ where P_i (respectively O_j) represents the i^{th} (respectively j^{th}) strategy of $S_P(x)$ (respectively S_O). The proponent can expect to get at least $\min_q E(x, p, q)$, where the minimum is taken over all the probability distributions q available to the opponent. Hence the proponent can choose a strategy which will guarantee her a reward of $\max_p \min_q E(x, p, q)$. The opposite is also true with $\min_q \max_p E(x, p, q)$.

Definition 4 (M&T semantics). *The semantics M&T is a gradual semantics that assigns a score to each argument $x \in \mathcal{A}$ in $\mathcal{F}$ as follows:*

$$Deg^{\mathtt{MT}}_{\mathcal{F}}(x) = \max_p \min_q E(x, p, q) = \min_q \max_p E(x, p, q)$$

3 Principles for Gradual Semantics

Principles have been introduced by [4] in order to better understand the behavior of the gradual semantics, choose a semantics for a particular application, guide the search for new semantics, compare semantics with each other, etc. We do not claim that all of these principles are mandatory (we will see later that some of them are incompatible). In the rest of this section, we introduce the principles.[6]

The first one, called Anonymity, states that the name of an argument should not impact its acceptability degree.

Principle 1 (Anonymity) *A semantics S satisfies Anonymity iff for any two AGs* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ *and* $\mathcal{F}' = (\mathcal{A}', \mathcal{R}')$ *for any isomorphism f from* $\mathcal{F}$ *to* $\mathcal{F}'$, $\forall a \in \mathcal{A}, Deg^S_{\mathcal{F}}(a) = Deg^S_{\mathcal{F}'}(f(a))$.

Independence says that the acceptability degree of an argument should be independent of unconnected arguments.

Principle 2 (Independence) *A semantics S satisfies Independence iff, for any two AGs* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ *and* $\mathcal{F}' = (\mathcal{A}', \mathcal{R}')$ *such that* $\mathcal{A} \cap \mathcal{A}' = \emptyset$, $\forall a \in \mathcal{A}$, $Deg^S_{\mathcal{F}}(a) = Deg^S_{\mathcal{F} \otimes \mathcal{F}'}(a)$.

Directionality states that the acceptability of argument x can depend on y only if there is a path from y to x.

Principle 3 (Directionality) *A semantics S satisfies Directionality iff, for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ *and* $\mathcal{F}' = (\mathcal{A}, \mathcal{R}')$ *such that* $a, b \in \mathcal{A}$, $\mathcal{R}' = \mathcal{R} \cup \{(a, b)\}$ *it holds that :* $\forall x \in \mathcal{A}$, *if there is no path from b to x, then* $Deg^S_{\mathcal{F}}(x) = Deg^S_{\mathcal{F}'}(x)$.

Neutrality states that an argument with an acceptability degree of 0 should have no impact on the arguments it attacks.

Principle 4 (Neutrality) *A semantics S satisfies Neutrality iff, for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ *if* $\forall a, b \in \mathcal{A}$, $\texttt{Att}_{\mathcal{F}}(b) = \texttt{Att}_{\mathcal{F}}(a) \cup \{x\}$ *with* $x \in \mathcal{A} \setminus \texttt{Att}_{\mathcal{F}}(a)$ *and* $Deg^S_{\mathcal{F}}(x) = 0$ *then* $Deg^S_{\mathcal{F}}(a) = Deg^S_{\mathcal{F}}(b)$.

Equivalence says that if two arguments have the same attackers, or more generally attackers of the same strength, they should have the same acceptability degree.

Principle 5 (Equivalence) *A semantics S satisfies Equivalence iff, for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, $\forall a, b \in \mathcal{A}$, *if there exists a bijective function f from* $\texttt{Att}_{\mathcal{F}}(a)$ *to* $\texttt{Att}_{\mathcal{F}}(b)$ *s.t.* $\forall x \in \texttt{Att}_{\mathcal{F}}(a)$, $Deg^S_{\mathcal{F}}(x) = Deg^S_{\mathcal{F}}(f(x))$ *then* $Deg^S_{\mathcal{F}}(a) = Deg^S_{\mathcal{F}}(b)$.

Maximality states that a non-attacked argument should have the highest acceptability degree.

Principle 6 (Maximality) *A semantics S satisfies Maximality iff, for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, $\forall a \in \mathcal{A}$, *if* $\texttt{Att}_{\mathcal{F}}(a) = \emptyset$ *then* $Deg^S_{\mathcal{F}}(a) = 1$.

[6] We do not include the Proportionality principle since it is only applicable when arguments are attached intrinsic weights.

Counting states that a non-zero degree attacker should impact the acceptability of the attacked argument.

Principle 7 (Counting) *A semantics S satisfies Counting iff for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, $\forall a, b \in \mathcal{A}$, *if i)* $Deg^S_{\mathcal{F}}(a) > 0$ *and ii)* $\texttt{Att}_{\mathcal{F}}(b) = \texttt{Att}_{\mathcal{F}}(a) \cup \{y\}$ *with* $y \in \mathcal{A} \backslash \texttt{Att}_{\mathcal{F}}(a)$ *and* $Deg^S_{\mathcal{F}}(y) > 0$ *then* $Deg^S_{\mathcal{F}}(a) > Deg^S_{\mathcal{F}}(b)$.

Weakening says that the acceptability of an argument should be strictly lower than 1 if it has at least one attacker with a non-zero acceptability degree.

Principle 8 (Weakening) *A semantics S satisfies Weakening iff for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, $\forall a \in \mathcal{A}$, *if* $\exists b \in \texttt{Att}_{\mathcal{F}}(a)$ *s.t.* $Deg^S_{\mathcal{F}}(b) > 0$, *then* $Deg^S_{\mathcal{F}}(a) < 1$.

Weakening Soundness states that if the acceptability degree of an argument is not maximal, it must be that it is attacked by at least one non-zero degree attacker.

Principle 9 (Weakening Soundness) *A semantics S satisfies Weakening Soundness iff, for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, $\forall a \in \mathcal{A}$, *if* $Deg^S_{\mathcal{F}}(a) < 1$ *then* $\exists b \in \texttt{Att}_{\mathcal{F}}(a)$ *such that* $Deg^S_{\mathcal{F}}(b) > 0$.

Reinforcement states that the acceptability degree increases if the acceptability degrees of attackers decrease.

Principle 10 (Reinforcement) *A semantics S satisfies Reinforcement iff for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, $\forall a, b \in \mathcal{A}$, *if i)* $Deg^S_{\mathcal{F}}(a) > 0$ *or* $Deg^S_{\mathcal{F}}(b) > 0$, *ii)* $\texttt{Att}_{\mathcal{F}}(a) \backslash \texttt{Att}_{\mathcal{F}}(b) = \{x\}$, *iii)* $\texttt{Att}_{\mathcal{F}}(b) \backslash \texttt{Att}_{\mathcal{F}}(a) = \{y\}$ *and iv)* $Deg^S_{\mathcal{F}}(y) > Deg^S_{\mathcal{F}}(x)$, *then* $Deg^S_{\mathcal{F}}(a) > Deg^S_{\mathcal{F}}(b)$.

Resilience states that no argument in an argumentation graph can have a acceptability degree of 0. It is certainly not a mandatory principle.

Principle 11 (Resilience) *A semantics S satisfies Resilience if for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, $\forall a \in \mathcal{A}$, $Deg^S_{\mathcal{F}}(a) > 0$.

The last three principles are incompatible with each other. The first principle, called Cardinality Precedence states, roughly speaking, that the greater the number of direct attackers of an argument, the lower its acceptability degree.

Principle 12 (Cardinality Precedence) *A semantics S satisfies Cardinality Precedence iff for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, $\forall a, b \in \mathcal{A}$, *if i)* $Deg^S_{\mathcal{F}}(b) > 0$, *and ii)* $|\{x \in \texttt{Att}_{\mathcal{F}}(a)$ *s.t.* $Deg^S_{\mathcal{F}}(x) > 0\}| > |\{y \in \texttt{Att}_{\mathcal{F}}(b)\ s.t.\ Deg^S_{\mathcal{F}}(y) > 0\}|$ *then* $Deg^S_{\mathcal{F}}(a) < Deg^S_{\mathcal{F}}(b)$.

Quality Precedence states, roughly speaking, that the greater the acceptability degree of the strongest attacker of an argument, the lower its acceptability degree.

Principle 13 (Quality Precedence) *A semantics S satisfies Quality Precedence if for any AG* $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, $\forall a, b \in \mathcal{A}$, *if i)* $Deg^S_{\mathcal{F}}(a) > 0$ *and ii)* $\exists y \in \texttt{Att}_{\mathcal{F}}(b)$ *s.t.* $\forall x \in \texttt{Att}_{\mathcal{F}}(a)$, $Deg^S_{\mathcal{F}}(y) > Deg^S_{\mathcal{F}}(x)$ *then* $Deg^S_{\mathcal{F}}(a) > Deg^S_{\mathcal{F}}(b)$.

Compensation states that several attacks from arguments with a low acceptability degree may compensate one attack from an argument with high acceptability degree. [7]

Principle 14 (Compensation) *A semantics S satisfies Compensation iff both Cardinality Precedence and Quality Precedence are not satisfied.*

In the literature, two principles directly refer to the self-attacking arguments. The first one, called Self-Contradiction, was introduced by [12] and states that the degree of a self-attacking argument should be strictly lower than the degree of an argument that does not attack itself.

Principle 15 (Self-Contradiction) *A semantics S satisfies Self-Contradiction iff, for any AG $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ with two arguments $a, b \in \mathcal{A}$, if $(a, a) \in \mathcal{R}$ and $(b, b) \notin \mathcal{R}$ then $Deg^S_{\mathcal{F}}(b) > Deg^S_{\mathcal{F}}(a)$.*

The second principle was introduced by Matt and Toni [19]. Its original name was "Self-contradiction must be avoided". We rename it for clarity reasons, namely in order to avoid the confusion with the name of Principle 15. This principle states that an argument that attacks itself should have the smallest acceptability degree (i.e. 0).

Principle 16 (Strong Self-Contradiction) *A semantics S satisfies Strong Self- Contradiction iff, for any AG $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ with $a \in \mathcal{A}$, $Deg^S_{\mathcal{F}}(a) = 0$ iff $(a, a) \in \mathcal{R}$.*

4 Analysis of Principles and Links Between Them

In this section we analyse the links between principles and identify two maximal mutually compatible sets of principles. Let us first observe that Strong Self-Contradiction implies Self-Contradiction. The next proposition follows directly from the definitions of the respective principles.

Proposition 1. *If a gradual semantics $\mathcal{S}$ satisfies Strong Self-Contradiction, it satisfies Self-Contradiction.*

Proof. Let us suppose that Strong Self-Contradiction is satisfied by $\mathcal{S}$. This means that those and only those arguments that have the minimum score are the self-attacking arguments ($\forall a \in \mathcal{A}$, $Deg^{\mathcal{S}}_{\mathcal{F}}(a) = 0$ iff $(a, a) \in \mathcal{R}$). This implies that all arguments that do not attack themselves have an acceptability degree greater than 0. Formally, $\forall b \in \mathcal{A}$, $Deg^{\mathcal{S}}_{\mathcal{F}}(b) > 0$ iff $(b, b) \notin \mathcal{R}$. Consequently, for two arguments $a, b \in \mathcal{A}$, if $(a, a) \in \mathcal{R}$ and $(b, b) \notin \mathcal{R}$ then $Deg^{\mathcal{S}}_{\mathcal{F}}(b) > Deg^{\mathcal{S}}_{\mathcal{F}}(a) = 0$. □

As discussed in the introduction, the next result shows that Equivalence and Self-Contradiction are incompatible.

Proposition 2. *There exists no gradual semantics $\mathcal{S}$ that satisfies both Equivalence and Self-Contradiction.*

[7] There are several version of this principle. We use the version that allows to clearly distinguish between the three cases (CP, QP, Compensation). Namely, each semantics satisfies *exactly* one of the three principles.

Proof. We provide a proof by contradiction. Let us suppose that a gradual semantics S satisfies both Equivalence and Self-Contradiction and consider the argumentation graph $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ with $\mathcal{A} = \{a, b\}$ and $\mathcal{R} = \{(a, a), (a, b)\}$.
From Self-Contradiction, we have $Deg^{\mathcal{S}}_{\mathcal{F}}(a) < Deg^{\mathcal{S}}_{\mathcal{F}}(b)$, while from Equivalence, we have $Deg^{\mathcal{S}}_{\mathcal{F}}(a) = Deg^{\mathcal{S}}_{\mathcal{F}}(b)$.
Contradiction. Hence, S does not satisfy both Equivalence and Self-Contradiction. Since S was arbitrary, we conclude that there exists no semantics that satisfies both Equivalence and Self-Contradiction. □

However, the Equivalence principle is not the only one incompatible with Strong Self-Contradiction. Some other incompatibilities exist mainly because self-attacking arguments are treated differently from other arguments. Indeed, according to Strong Self-Contradiction, self-attacking arguments are directly classified as the worst arguments, whereas the other principles just consider a self-attack as an attack like any other (i.e. an attack between two distinct arguments).

Proposition 3. *There exists no gradual semantics $\mathcal{S}$ that satisfies both Strong Self-Contradiction and Resilience.*

Proof. We provide a proof by contradiction. Let us suppose that a gradual semantics S satisfies both Strong Self-Contradiction and Resilience, and consider the argumentation graph $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ where $\mathcal{A} = \{a\}$ and $R = \{(a, a)\}$.
From Strong Self-Contradiction, we have $Deg^{\mathcal{S}}_{\mathcal{F}}(a) = 0$, while from Resilience, we have $Deg^{\mathcal{S}}_{\mathcal{F}}(a) > 0$.
Contradiction. Hence, S does not satisfy both Strong Self-Contradiction and Resilience. Since S was arbitrary, there exists no semantics that satisfies both Resilience and Strong Self-Contradiction. □

Proposition 4. *There exists no gradual semantics $\mathcal{S}$ that satisfies both Strong Self-Contradiction and Weakening Soundness.*

Proof. We provide a proof by contradiction. Let us suppose that a gradual semantics S satisfies both Strong Self-Contradiction and Weakening Soundness, and consider the argumentation graph $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ where $\mathcal{A} = \{a\}$ and $R = \{(a, a)\}$.
From Strong Self-Contradiction, we have $Deg^{\mathcal{S}}_{\mathcal{F}}(a) = 0$, while from Weakening Soundness, we have $Deg^{\mathcal{S}}_{\mathcal{F}}(a) > 0$ because a is the only attacker of a and $Deg^{\mathcal{S}}_{\mathcal{F}}(a) = 0$.
Contradiction. Hence, S does not satisfy both Strong Self-Contradiction and Weakening Soundness. Since S was arbitrary, there exists no semantics that satisfies both Strong Self-Contradiction and Weakening Soundness. □

Proposition 5. *There exists no gradual semantics $\mathcal{S}$ that satisfies both Strong Self-Contradiction and Reinforcement.*

Proof. We provide a proof by contradiction. Let us suppose that a gradual semantics S satisfies both Strong Self-Contradiction and Reinforcement, and consider the argumentation graph $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ represented in Figure 1.

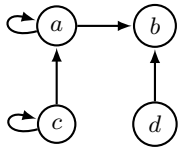


Fig. 1: AG showing that Reinforcement and Strong Self-Contradiction are incompatible.

From Strong Self-Contradiction, we have $0 = Deg^{\mathcal{S}}_{\mathcal{F}}(a) < Deg^{\mathcal{S}}_{\mathcal{F}}(b)$. From Reinforcement, we have $Deg^{\mathcal{S}}_{\mathcal{F}}(a) > Deg^{\mathcal{S}}_{\mathcal{F}}(b)$ because i) $Deg^{\mathcal{S}}_{\mathcal{F}}(b) > 0$, ii) $\texttt{Att}_{\mathcal{F}}(a) \backslash \texttt{Att}_{\mathcal{F}}(b) = \{c\}$, iii) $\texttt{Att}_{\mathcal{F}}(b) \backslash \texttt{Att}_{\mathcal{F}}(a) = \{d\}$, and iv) $Deg^{\mathcal{S}}_{\mathcal{F}}(d) > Deg^{\mathcal{S}}_{\mathcal{F}}(c)$.
Contradiction. Hence, S does not satisfy both Strong Self-Contradiction and Reinforcement. Since S was arbitrary, there exists no semantics that satisfies both Strong Self-Contradiction and Reinforcement. □

Proposition 6. *There exists no gradual semantics $\mathcal{S}$ that satisfies both Strong Self-Contradiction and Neutrality.*

Proof. We provide a proof by contradiction. Let us suppose that a gradual semantics S satisfies both Strong Self-Contradiction and Neutrality, and consider the argumentation graph $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ represented in Figure 2.

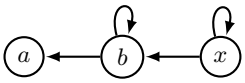


Fig. 2: AG showing that Neutrality and Strong Self-Contradiction are incompatible.

From Strong Self-Contradiction, we have $0 = Deg^{\mathcal{S}}_{\mathcal{F}}(b) < Deg^{\mathcal{S}}_{\mathcal{F}}(a)$. From Neutrality, we have $Deg^{\mathcal{S}}_{\mathcal{F}}(a) = Deg^{\mathcal{S}}_{\mathcal{F}}(b)$ because $\texttt{Att}_{\mathcal{F}}(b) = \texttt{Att}_{\mathcal{F}}(a) \cup \{x\}$ with $Deg^{\mathcal{S}}_{\mathcal{F}}(x) = 0$.
Contradiction. Hence, S does not satisfy both Strong Self-Contradiction and Neutrality. Since S was arbitrary, there exists no semantics that satisfies both Strong Self-Contradiction and Neutrality. □

Taking these incompatibilities into account, our goal is now to study two maximal mutually compatible sets of principles we are interested in. For this, we need the notion of dominance. A semantics S dominates a semantics S' on the set of principles P if the subset of principles from P satisfied by S is a strict superset of the subset of principles from P satisfied by S'. In the rest of the discussion, we suppose that P is the set of all principles studied in Section 3. Note that if a semantics S satisfies a maximal for set inclusion set of principles, it is not dominated by any semantics.

A first maximal (for set inclusion) set of principles has been identified by [4] and is a direct consequence of their Proposition 1. We define this set of principles as $P_{CREW} =$

{Anonymity, Independence, Directionality, Neutrality, Equivalence, Maximality, Weakening, Counting, Weakening Soundness, Reinforcement, Resilience and Compensation}.

Theorem 1 ([4]). *P_{CREW} is a maximal for set inclusion set of principles.*

We can formally show that there is a unique maximal set of principles compatible with Compensation, Resilience, Equivalence and Weakening Soundness.

Theorem 2. *Let P be the set of all principles defined in Section 3 (Principles 1-16). Let $\mathcal{S}$ be a gradual semantics that satisfies Compensation, Resilience, Equivalence and Weakening Soundness. If $\mathcal{S}$ is not dominated w.r.t. P, then $\mathcal{S}$ satisfies exactly the principles from P_{CREW}.*

Proof. On one hand, we know from the work by [4] that h-categorizer satisfies all the principles from P_{CREW}. On the other hand, it is clear from the incompatibility results between the principles that $\mathcal{S}$ cannot satisfy Strong Self-Contradiction which is incompatible with Resilience (see Proposition 3), Self-Contradiction which is incompatible with Equivalence (see Proposition 2), Cardinality/Quality Precedence which are both incompatible with Compensation (see [4]). Thus, in order not to be dominated by h-categorizer, $\mathcal{S}$ must satisfy all the principles from P_{CREW}; due to the incompatibilities, $\mathcal{S}$ cannot satisfy any more principles. □

In this paper we choose to explore the space of principles compatible with Strong Self-Contradiction (which is not in P_{CREW}). One naturally wants to maximise the set of satisfied principles. Can we satisfy Strong Self-Contradiction and all the other principles? The answer is negative (see Propositions 2-6). First, one has to choose between Cardinality Precedence, Quality Precedence and Compensation. In this paper, we explore the possibility of satisfying Compensation. This choice is based on the fact that this principle is satisfied by virtually all semantics, as showed by Amgoud et al. [4]. Indeed, Cardinality Precedence and Quality Precedence represent, roughly speaking, *drastic* or *extreme* cases and are satisfied only by the semantics specifically designed to satisfy them, like max-based semantics and card-based semantics [4] or by semantics having other specificities. For instance, iterative schema [18], which satisfies Quality Precedence, is a discrete semantics (it takes only three possible values). This yields another maximal set of principles which includes those two principles. We define this set of principles as P_{2S2C} = {Anonymity, Independence, Directionality, Maximality, Weakening, Counting, Compensation, Self-Contradiction, Strong Self-Contradiction}.

Theorem 3. *P_{2S2C} is a maximal for set inclusion set of principles.*

Proof. Firstly, all the principles in P_{2S2C} are compatible because `nsa` satisfies all of them (see Proposition 7). Secondly, P_{2S2C} is maximal because for each remaining principle $p \in$ {Equivalence, Weakening Soundness, Neutrality, Reinforcement, Cardinality Precedence, Quality Precedence and Resilience}, there exists (at least) one principle in P_{2S2C} which is incompatible with p:

- Equivalence and Self-Contradiction are incompatible (see Proposition 2);
- Neutrality and Strong Self-Contradiction are incompatible (see Proposition 6);

- Reinforcement and Strong Self-Contradiction are incompatible (see Proposition 5);
- Weakening Soundness and Strong Self-Contradiction are incompatible (see Proposition 4);
- Cardinality Precedence and Compensation are incompatible (see [4]);
- Quality Precedence and Compensation are incompatible (see [4]);
- Resilience and Strong Self-Contradiction are incompatible (see Proposition 3);

□

We now show that there is a unique maximal set of principles compatible with Strong Self-Contradiction and Compensation. This follows from the fact that if a semantics satisfies Strong Self-Contradiction, it cannot satisfy several principles (see Propositions 2-6) but can satisfied all the others (as witnessed by the semantics we introduce in this paper).

Theorem 4. *Let P be the set of all principles defined in Section 3 (Principles 1-16). Let $\mathcal{S}$ be a gradual semantics that satisfies Strong Self-Contradiction and Compensation. If $\mathcal{S}$ is not dominated w.r.t. P, then $\mathcal{S}$ satisfies exactly the principles from P_{2S2C}.*

Proof. It is clear that from the incompatibility results between different principles, $\mathcal{S}$ cannot satisfy (i) Resilience, Equivalence and Weakening Soundness which are incompatible with Strong Self-Contradiction (or Self-Contradiction), and (ii) Cardinality Precedence and Quality Precedence which are both incompatible with Compensation. The set of remaining principles corresponds exactly to P_{2S2C} which is a maximal for set inclusion set of principles. However, $\mathcal{S}$ cannot satisfy exactly a subset of P_{2S2C} because, in this case, $\mathcal{S}$ will be dominated by a semantics that satisfies the principles of P_{2S2C}. Consequently, when $\mathcal{S}$ satisfies Strong Self-Contradiction and Compensation, the only way to ensure that $\mathcal{S}$ is not dominated is when $\mathcal{S}$ satisfies exactly the principles from P_{2S2C}. □

To the best of our knowledge, no semantics that satisfy all the principles from P_{2S2C} has been presented in the literature. In the next section, we define a semantics that satisfies this set of principles.

Before doing that, let us comment on the non satisfaction of some principles. It is tempting to change the principles in order to treat the self-attacks in another way, and consequently make the principles fit some definitions or theorems. We argue that it is better to start by having a full picture of what happens with *existing* principles. Indeed, the principles should be the most stable part of a theory. We are not against introduction of new principles (or changing the existing ones). This might be part of future work.

5 No Self-Attack h-categorizer Semantics

In this section, we define a new gradual semantics, called no self-attack h-categorizer (`nsa`) semantics, inspired by the h-categorizer semantics. The main difference is that we assign 0 degree to the self-attacking arguments.

Definition 5 (nsa). *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AG. We define $f_{\texttt{nsa}}^{\mathcal{F},i} : \mathcal{A} \rightarrow [0, +\infty]$ as follows : for every argument $a \in \mathcal{A}$ for $i \in \{0, 1, 2, ..\}$,*

$$f_{\texttt{nsa}}^{\mathcal{F},i}(a) = \begin{cases} 0 & \text{if } (a,a) \in \mathcal{R} \\ 1 & \text{if } (a,a) \notin \mathcal{R} \text{ and } i = 0 \\ \dfrac{1}{1 + \sum_{b \in \texttt{Att}_{\mathcal{F}}(a)} f_{\texttt{nsa}}^{\mathcal{F},i-1}(b)} & \text{if } (a,a) \notin \mathcal{R} \text{ and } i > 0 \end{cases} \tag{1}$$

By convention, if $\texttt{Att}_{\mathcal{F}}(a) = \emptyset$, $\sum_{b \in \texttt{Att}_{\mathcal{F}}(a)} f_{\texttt{nsa}}^{\mathcal{F},i-1}(b) = 0$.

Although nsa is inspired by the h-categorizer semantics, the modifications made change the result obtained requiring the verification that nsa also converges to a unique result. Thus, in the next result, we show that for every argumentation graph $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, for every argument $a \in \mathcal{A}$, $f_{\texttt{nsa}}^{\mathcal{F},i}(a)$ converges as i approaches infinity. Roughly speaking, the goal of the next theorem is to formally check that assigning zero values to self-attacking arguments does not impact the convergence of the scores. Thus, applying nsa to the original argumentation graph $\mathcal{F}$ provides the same result as when the h-categorizer semantics is applied on a restricted version of $\mathcal{F}$ where the self-attacking arguments are deleted.

Theorem 5. *For every argumentation graph $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, for every $a \in \mathcal{A}$, if $(a,a) \notin \mathcal{R}$, we have $\lim_{i \to \infty} f_{\texttt{nsa}}^{\mathcal{F},i}(a) = Deg_{\mathcal{F}'}^{\texttt{h}}(a)$ where $\mathcal{F}' = (\mathcal{A}', \mathcal{R}')$ with $\mathcal{A}' = \{x \in \mathcal{A} | \, (x,x) \notin \mathcal{R}\}$ and $\mathcal{R}' = \{(x,y) \in \mathcal{R} \mid x \in A' \text{ and } y \in A'\}$.*

Proof. Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AG and $\mathcal{F}' = (\mathcal{A}', \mathcal{R}')$ be an AG such that $\mathcal{A}' = \{x \in \mathcal{A} | (x,x) \notin \mathcal{R}\}$ and $\mathcal{R}' = \{(x,y) \in \mathcal{R} \mid x \in A' \text{ and } y \in A'\}$. Without loss of generality, let us denote $\mathcal{A} = \{a_0, a_1, \ldots, a_n\}$.

Let us recall the iterative version of h-categorizer, that can be used to calculate the scores of arguments [20]: for every a, for $i \in \mathbb{N}$

$$f_h^{\mathcal{F},i}(a) = \begin{cases} 1 & \text{if } i = 0 \\ \dfrac{1}{1 + \sum_{b \in \texttt{Att}_{\mathcal{F}}(a)} f_h^{\mathcal{F},i-1}(b)} & \text{if } i > 0 \end{cases} \tag{2}$$

We prove by induction on i that for each $a \in \mathcal{A}'$:

$$f_{\texttt{nsa}}^{\mathcal{F},i}(a) = f_h^{\mathcal{F}',i}(a)$$

Base: Let $i = 0$. From the formal definition of nsa (Definition 5) and equation (2), we have $f_{\texttt{nsa}}^{\mathcal{F},0}(a) = f_h^{\mathcal{F}',0}(a) = 1$. Thus, the inductive base holds.

Step: Let us suppose that the inductive hypothesis is true for every $k \in \{0, 1, \ldots i\}$ and let us show that it is true for $i + 1$. We need to prove :

$$f_{\texttt{nsa}}^{\mathcal{F},i+1}(a) = f_h^{\mathcal{F}',i+1}(a)$$

From the inductive hypothesis, we know that for each argument $a \in A'$, $f_{\texttt{nsa}}^{\mathcal{F},i}(a) = f_h^{\mathcal{F}',i}(a)$. Thus, from equation (1), we have:

$$f_{\texttt{nsa}}^{\mathcal{F},i+1}(a) = \frac{1}{1 + \sum_{b \in \texttt{Att}_{\mathcal{F}}(a)} f_{\texttt{nsa}}^{\mathcal{F},i}(b)}$$

From equation (2), we have

$$f_h^{\mathcal{F}',i+1}(a) = \frac{1}{1 + \sum_{b \in \texttt{Att}_{\mathcal{F}'}(a)} f_h^{\mathcal{F}',i}(b)}$$

Let us note $\texttt{Att}_{\mathcal{F}}(a) = \texttt{Att}_{\mathcal{F}'}(a) \cup \{b_0, \ldots, b_m\}$ with $m \geq 0$ and remark that $\forall b \in \{b_0, \ldots, b_m\}$, we have $(b, b) \in \mathcal{R}$. According to equation (1), $\forall b \in \{b_0, \ldots, b_m\}$, $f_{\texttt{nsa}}^{\mathcal{F},i}(b) = 0$. Consequently, as 0 is the neutral element of the addition, we have $\forall a \in \mathcal{A}'$, $f_{\texttt{nsa}}^{\mathcal{F},i+1}(a) = f_h^{\mathcal{F}',i+1}(a)$.

By induction, we conclude that for every $i \in \mathbb{N}$ and for every $a \in \mathcal{A}'$

$$f_{\texttt{nsa}}^{\mathcal{F},i}(a) = f_h^{\mathcal{F}',i}(a)$$

Since f_h converges when $i \to \infty$ and $f_{\texttt{nsa}}$ coincides with f_h for every argument of $\mathcal{A}'$, we conclude that $f_{\texttt{nsa}}$ converges too. Formally, $\forall a \in A'$,

$$\lim_{i \to \infty} f_{\texttt{nsa}}^{\mathcal{F},i}(a) = \lim_{i \to \infty} f_{\texttt{h}}^{\mathcal{F},i}(a) = Deg_{\mathcal{F}'}^{\texttt{h}}(a)$$

□

We can now introduce the formal definition of nsa.

Definition 6 (nsa). *The no self-attack h-categorizer semantics is a function* nsa *which associates to any argumentation framework $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ a function $Deg_{\mathcal{F}}^{\texttt{nsa}}(a) : A \to [0, 1]$ as follows: $Deg_{\mathcal{F}}^{\texttt{nsa}}(a) = \lim_{i \to \infty} f_{nsa}^{\mathcal{F},i}(a)$.*

We can now show that the acceptability degrees attributed to arguments by nsa satisfy the equation from Definition 5 (naturally, not taking into account the second line of the equation, since it considers the case $i = 0$).

Theorem 6. *For any $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, for any $a \in \mathcal{A}$,*

$$Deg_{\mathcal{F}}^{\texttt{nsa}}(a) = \begin{cases} 0 & \text{if } (a, a) \in \mathcal{R} \\ \dfrac{1}{1 + \sum_{b \in \texttt{Att}_{\mathcal{F}}(a)} Deg_{\mathcal{F}}^{\texttt{nsa}}(b)} & \text{otherwise} \end{cases}$$

Proof. Let $\mathcal{F} = (A, R)$ be an argumentation graph and $a \in A$.
The case where a is a self-attacking argument is trivial.
In the rest of the proof we consider the case where a is not a self-attacking argument. Letting $\lim_{i \to \infty}$ in the following equality

$$f_{\texttt{nsa}}^{i+1}(a) = \frac{1}{1 + \sum_{b \in \texttt{Att}_{\mathcal{F}}(a)} f_{\texttt{nsa}}^{i}(b)}$$

and using the fact that arithmetical operations and sum are continuous functions, we obtain :

$$\lim_{i\to\infty} f^{i+1}_{\mathtt{nsa}}(a) = \frac{1}{1+\sum_{b\in \mathtt{Att}_{\mathcal{F}}(a)} \lim\limits_{i\to\infty} f^{i}_{\mathtt{nsa}}(b)}$$

then

$$Deg^{\mathtt{nsa}}_{\mathcal{F}}(a) = \frac{1}{1+\sum_{b\in \mathtt{Att}_{\mathcal{F}}(a)} Deg^{\mathtt{nsa}}_{\mathcal{F}}(a)}$$

□

We now show that the equation from Theorem 6 is not only satisfied by `nsa`, but is also its characterization. More precisely, the next result proves that if an arbitrary semantics D satisfies that equation, it must be that D coincides with `nsa`.

Theorem 7. *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AG with $a \in \mathcal{A}$ and $D : \mathcal{A} \to [0,1]$ be a function with the following formula:*

$$D(a) = \begin{cases} 0 & \text{if } (a,a) \in \mathcal{R} \\ \dfrac{1}{1+\sum_{b\in Att_{\mathcal{F}}(a)} D(b)} & \text{otherwise} \end{cases} \tag{3}$$

then $D \equiv Deg^{\mathtt{nsa}}_{\mathcal{F}}$.

Proof. Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AG and suppose that $D : \mathcal{A} \to [0,1]$ is the function from equation (3).

Let $A = \{a_1, .., a_n\}$ and let $F : [0,1]^n \to [0,1]^n$ be the function such that $F(x_1, .., x_n) = (F_1(x_1, .., x_n), ..., F_n(x_1, ..., x_n))$ where the functions F_i are defined by the following equality:

$$F_i(x_1, \ldots, x_n) = \begin{cases} 0 & \text{if } (a_i, a_i) \in \mathcal{R} \\ \dfrac{1}{1+\sum\limits_{j:a_j\in Att_{\mathcal{F}}(a_i)} x_j} & \text{otherwise} \end{cases} \tag{4}$$

We also define the partial order $\leq$ on $\mathbb{R}^n$ in the following way: if $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ then $x \leq y$ iff for every i it holds that $x_i \leq y_i$.

Thus, from Equation (3), it follows that

$$F(D(a_1), ..., D(a_n)) = (D(a_1), ..., D(a_n)).$$

Observe that F is a non-increasing function and that $G = F \circ F$ is a non-decreasing function, and that :

$$(f^{i+1}_{\mathtt{nsa}}(a_1), ..., f^{i+1}_{\mathtt{nsa}}(a_n)) = F((f^{i}_{\mathtt{nsa}}(a_1), ..., f^{i}_{\mathtt{nsa}}(a_n)))$$

for every $i \in \mathbb{N}$. Since $(f^0_{\mathtt{nsa}}(a_1), ..., f^0_{\mathtt{nsa}}(a_n)) \in [0,1]^n$ with $f^0_{\mathtt{nsa}}(a_i) = 0$ iff $(a_i, a_i) \in \mathcal{R}$ and $f^0_{\mathtt{nsa}}(a_i) = 1$ otherwise, by the inequalities, we obtain

$$(f^0_{\mathtt{nsa}}(a_1), ..., f^0_{\mathtt{nsa}}(a_n)) \geq (D(a_1), ..., D(a_n)) \tag{5}$$

From (5), and since F is non-increasing, we have:

$$(f^1_{\texttt{nsa}}(a_1), ..., f^1_{\texttt{nsa}}(a_n)) \leq (D(a_1), ..., D(a_n)) \tag{6}$$

From (6), and since $G = F \circ F$ is non-decreasing, we have:

$$(f^{2i}_{\texttt{nsa}}(a_1), ..., f^{2i}_{\texttt{nsa}}(a_n)) \geq (D(a_1), ..., D(a_n)) \tag{7}$$

and

$$(f^{2i+1}_{\texttt{nsa}}(a_1), ..., f^{2i+1}_{\texttt{nsa}}(a_n)) \leq (D(a_1), ..., D(a_n)) \tag{8}$$

for every $i \in \mathbb{N}$.
Since all f^i converge, from (7) and (8) we obtain

$$(Deg^{\texttt{nsa}}_{\mathcal{F}}(a_1), \ldots, Deg^{\texttt{nsa}}_{\mathcal{F}}(a_n)) \geq (D(a_1), ..., D(a_n))$$

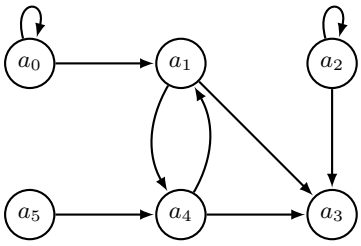
and

$$(Deg^{\texttt{nsa}}_{\mathcal{F}}(a_1), \ldots, Deg^{\texttt{nsa}}_{\mathcal{F}}(a_n)) \leq (D(a_1), ..., D(a_n))$$

and thus $\forall a \in \mathcal{A}, Deg^{\texttt{nsa}}_{\mathcal{F}}(a) = D(a)$. □

Below is an example of the nsa semantics applied on an argumentation graph.

Example 1 *Let us apply the no self-attack h-categorizer semantics (*nsa*) on the argumentation graph illustrated in Fig. 3. By definition, the self-attacking arguments have*



$Deg^{\mathcal{S}}_{\mathcal{F}}$	nsa	h	MT
a_0	0	0.618	0
a_1	0.732	0.495	0.25
a_2	0	0.618	0
a_3	0.477	0.398	0.167
a_4	0.399	0.401	0.25
a_5	1	1	1

Fig. 3: On the left, an argumentation graph $\mathcal{F}$ and, on the right, the table containing the degrees of acceptability of each argument of $\mathcal{F}$ w.r.t. the no self-attack h-categorizer semantics (nsa), the h-categorizer semantics (h) and the semantics M&T (MT).

an acceptability degree of 0 : $Deg^{\texttt{nsa}}_{\mathcal{F}}(a_0) = Deg^{\texttt{nsa}}_{\mathcal{F}}(a_2) = 0$*. The non-attacked arguments or the arguments only attacked by self-attacking arguments have, by definition, the maximum score:* $Deg^{\texttt{nsa}}_{\mathcal{F}}(a_5) = 1$*. Applying the formula from Theorem 6, we obtain the following acceptability degrees for* a_1 *and* a_4 *:* $Deg^{\texttt{nsa}}_{\mathcal{F}}(a_1) = 0.732$ *and* $Deg^{\texttt{nsa}}_{\mathcal{F}}(a_4) = 0.399$*. Finally, following the same method, here are the details concerning* a_3 *:*

$$\begin{aligned} Deg^{\texttt{nsa}}_{\mathcal{F}}(a_3) &= \frac{1}{1 + Deg^{\texttt{nsa}}_{\mathcal{F}}(a_1) + Deg^{\texttt{nsa}}_{\mathcal{F}}(a_2) + Deg^{\texttt{nsa}}_{\mathcal{F}}(a_4)} \\ &= \frac{1}{1 + 0.732 + 0 + 0.399} \\ &= 0.477 \end{aligned}$$

In order to have an overview of the difference between nsa *and the gradual semantics introduced in Section 2, the degrees of acceptability of arguments w.r.t. the h-categorizer semantics and the M&T semantics have also been added in the table of Fig. 3. This comparison clearly shows that nullifying the impact of self-attacking arguments more or less significantly changes the degree of acceptability of other arguments (e.g.* a_1 *and* a_3*).*

6 Principle-Based Evaluation of Semantics

In this section we evaluate the nsa semantics with respect to principle compliance, and compare the results with two existing semantics, namely M&T and h-categorizer. We first show that nsa satisfies all the principles from P_{2S2C}, and thus cannot be dominated by any semantics.

Proposition 7. *The gradual semantics* nsa *satisfies all the principles from* P_{2S2C}*. The other principles are not satisfied.*

In order to axiomatically compare nsa with the two other gradual semantics, let us check for the principles studied in this paper those that are satisfied by M&T and recall those satisfied by the h-categorizer semantics.

Proposition 8. *The gradual semantics M&T satisfies Anonymity, Independence, Directionality, Maximality, Weakening, Compensation, Self-Contradiction and Strong Self-Contradiction. The other principles are not satisfied.*

Proposition 9 ([3]). *The gradual semantics* h*-categorizer satisfies all the principles from* P_{CREW}*. The other principles are not satisfied.*

Note that nsa dominates M&T, i.e. it satisfies strictly more principles. Observe that nsa and h-categorizer are incomparable in terms of principles satisfaction. Indeed, nsa represents one choice, i.e. the position to satisfy Strong Self-Contradiction and Compensation. It also satisfies all the compatible principles. h-categorizer represents another choice, namely that to satisfy Compensation, Resilience, Equivalence and Weakening Soundness. Concretely, a semantics satisfying P_{CREW} considers that a self-attacking argument is a path like the other ones. So an argument which attacks itself (and is not attacked by any other argument) can be stronger than an argument which is attacked by several arguments. On the contrary, a semantics which satisfies P_{2S2C} considers that a self-attacking argument is intrinsically flawed, without even requiring other arguments to defeat it. Note that there exist other maximal sets of compatible principles, for example the one containing Resilience and Self-Contradiction. We leave a detailed study of these maximal sets of compatible principles for future work.

7 Experimental Results

We now empirically compare nsa with M&T and h-categoriser semantics. We consider a large experimental setting representing three different models used during the

Principles	M&T	h-cat	nsa
Anonymity	✓	✓	✓
Independence	✓	✓	✓
Directionality	✓	✓	✓
Neutrality	×	✓	×
Equivalence	×	✓	×
Maximality	✓	✓	✓
Weakening	✓	✓	✓
Counting	×	✓	✓
Weakening Soundness	×	✓	×
Reinforcement	×	✓	×
Resilience	×	✓	×
Cardinality Precedence	×	×	×
Quality Precedence	×	×	×
Compensation	✓	✓	✓
Self-Contradiction	✓	×	✓
Strong Self-Contradiction	✓	×	✓

Table 1: Principles satisfied by the M&T, h-categorizer and `nsa` semantics. The shaded cells contain the results already proved in the literature.

ICCMA competition (http://argumentationcompetition.org/) as a way to generate random argumentation graphs: i) the Erdös-Rényi model (ER) which generates graphs by randomly selecting attacks between arguments, ii) the Barabasi-Albert model (BA) which provides networks, called scale-free networks, with a structure in which some nodes have a huge number of links, but in which nearly all nodes are connected to only a few other nodes, and iii) the Watts-Strogatz model (WS) which produces graphs which have small-world network properties, such as high clustering and short average path lengths. The generation of these three types of AGs was done by the AFBench-Gen2 generator [15]. We generated a total of 2160 AGs evenly distributed between the three models. For each model, the number of arguments varies among $Arg = \{5, 10, 15, 25, 50, 100, 250, 500\}$ with 90 AGs for each of these values. The parameters used to generate graphs are as follows: for ER, 10 random instances for each (numArg, probAttacks) in $Arg \times \{0.2, 0.3, \ldots, 1\}$; for BA, 9 random instances for each (numArg, probCycles) in $Arg \times \{0, 0.1, \ldots, 0.9\}$; for WS, (numArg, probCycles, β, $\mathcal{K}$) in $Arg \times \{0.25, 0.5, 0.75\} \times \{0, 0.25, 0.5, 0.75, 1\} \times \{k \in 2\mathbb{N}\ s.t.\ 2 \leq k \leq |Arg| - 1\}$. We refer the reader to [15] for the meaning of the parameters.

In order to compare the execution times of the three semantics studied in this paper, we have implemented them in C and ran the program on a cluster of identical computers with dual quad-core processors with 128 GB RAM.[8]

Figure 4 shows the average execution time obtained by each semantics for the instances classified according to the number of arguments. A first remark is that, unlike

[8] The code and benchmarks are available online at https://github.com/jeris90/nsa_code.git.

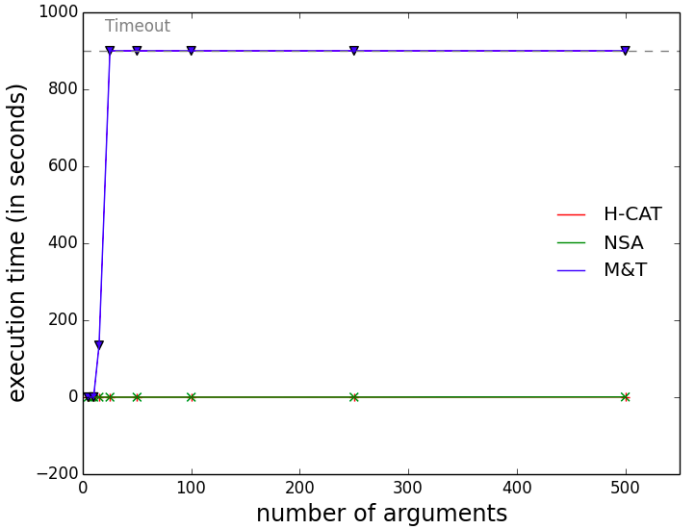


Fig. 4: Execution speed for the `nsa` (in green), the M&T (in blue) and the h-categorizer (in red) semantics. x-axis shows the number of arguments of the instances ($Arg = \{5, 10, 15, 25, 50, 100, 250, 500\}$). y-axis shows the execution time in seconds (with a timeout of 900 seconds).

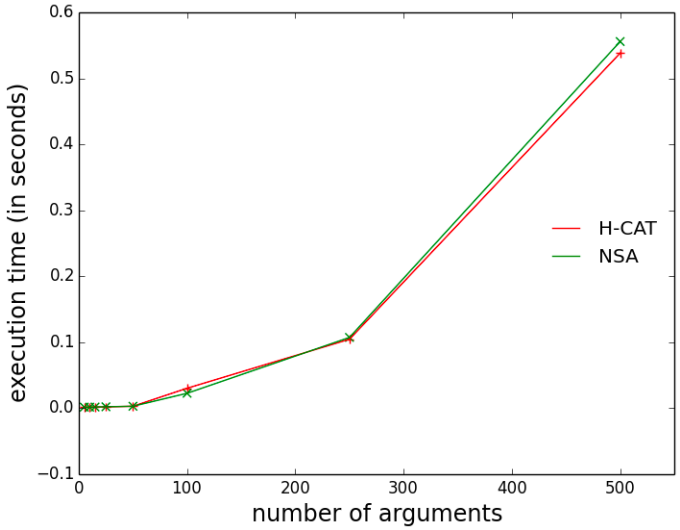


Fig. 5: A zoomed-in version of the graph from Figure 4 to better see the difference between the execution speed for the `nsa` semantics (in green) and the h-categorizer (in red) semantics.

the other two semantics, the M&T semantics quickly explodes in time since it systematically reaches the timeout (900 seconds) when the number of arguments is greater than 15. A second remark is that, unsurprisingly, the `nsa` and h-categorizer semantics have very similar execution times for each of the instances. Figure 5 shows the difference between `nsa` and h-categorizer semantics more precisely. Moreover, they allow us to quickly compute (with an average smaller than one second) the degree of acceptability of each argument even for large AGs. Only a few very dense instances (i.e. those with a high probability of cycles) require between 1 and 2 seconds when $\mathsf{numArg} = 500$.

8 Summary

We studied the question of the treatment of self-attacks by gradual semantics following a principle-based approach. We showed links and incompatibilities between principles, defined a new semantics called no self-attack h-categorizer semantics and proved that it dominates the only existing semantics satisfying Self-Contradiction principle. Moreover, we showed that our semantics satisfies a maximal possible amount of principles (i.e. no semantics satisfying Self-Contradiction can satisfy more principles) and is usable in practice as it returns results very quickly (on average less than 1 second) even on large and dense AGs.
In addition to the future work already discussed in the paper, we think it would be interesting to extend the approach we used for the h-categorizer semantics to other gradual semantics (if possible). Finally, the work presented in this paper concerns "classic" argumentation graphs but one could naturally ask the same question about AGs containing more information (support relation, weight on arguments and/or attacks, etc.).

9 Acknowledgements

Vivien Beuselinck was supported by the ANR-3IA Artificial and Natural Intelligence Toulouse Institute. Srdjan Vesic was supported by Responsible AI Chair a chair in Artificial Intelligence (https://ia-responsable.eu/).

References

1. Amgoud, L.: A replication study of semantics in argumentation. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19) (2019)
2. Amgoud, L., Ben-Naim, J.: Ranking-based semantics for argumentation frameworks. In: Proc. of the 7th International Conference on Scalable Uncertainty Management (SUM'13). pp. 134–147 (2013)
3. Amgoud, L., Ben-Naim, J.: Axiomatic foundations of acceptability semantics. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016. pp. 2–11. AAAI Press (2016)
4. Amgoud, L., Ben-Naim, J., Doder, D., Vesic, S.: Acceptability semantics for weighted argumentation frameworks. In: Sierra, C. (ed.) Proc. of the 26th International Joint Conference on Artificial Intelligence, (IJCAI'17). pp. 56–62 (2017)

5. Amgoud, L., Doder, D.: Gradual semantics accounting for varied-strength attacks. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, (AAMAS'19). pp. 1270–1278 (2019)
6. Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. The Knowledge Engineering Review **26**(4), 365–410 (2011)
7. Baroni, P., Giacomin, M.: Solving semantic problems with odd-length cycles in argumentation. In: Proc. of the 7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, (ECSQARU'03). vol. 2711, pp. 440–451 (2003)
8. Baumann, R., Brewka, G., Ulbricht, M.: Comparing weak admissibility semantics to their Dung-style counterparts - reduct, modularization, and strong equivalence in abstract argumentation. In: Calvanese, D., Erdem, E., Thielscher, M. (eds.) Proc. of the 17th International Conference on Principles of Knowledge Representation and Reasoning, (KR'20). pp. 79–88 (2020)
9. Baumann, R., Brewka, G., Ulbricht, M.: Revisiting the foundations of abstract argumentation - semantics based on weak admissibility and weak defense. In: Proc. of the 34th AAAI Conference on Artificial Intelligence, (AAAI'20). pp. 2742–2749 (2020)
10. Besnard, P., Hunter, A.: A logic-based theory of deductive arguments. Artificial Intelligence **128**(1-2), 203–235 (2001)
11. Bodanza, G.A., Tohmé, F.A.: Two approaches to the problems of self-attacking arguments and general odd-length cycles of attack. J. Appl. Log. **7**(4), 403–420 (2009)
12. Bonzon, E., Delobelle, J., Konieczny, S., Maudet, N.: A Comparative Study of Ranking-based Semantics for Abstract Argumentation. In: Proc. of the 30th AAAI Conference on Artificial Intelligence (AAAI'16). pp. 914–920 (2016)
13. Bonzon, E., Delobelle, J., Konieczny, S., Maudet, N.: Combining extension-based semantics and ranking-based semantics for abstract argumentation. In: Proc. of the 16th International Conference on Principles of Knowledge Representation and Reasoning (KR'18). pp. 118–127 (2018)
14. Caminada, M.: On the issue of reinstatement in argumentation. In: Proc. of the 10th European Conference on Logics in Artificial Intelligence (JELIA'06). pp. 111–123 (2006)
15. Cerutti, F., Vallati, M., Giacomin, M.: Afbenchgen2: A generator for random argumentation frameworks (2017), http://argumentationcompetition.org/2017/AFBenchGen2.pdf
16. Dauphin, J., Rienstra, T., van der Torre, L.: A principle-based analysis of weakly admissible semantics. In: Proc. of the 8th International Conference on Computational Models of Argument, (COMMA'20). Frontiers in Artificial Intelligence and Applications, vol. 326, pp. 167–178 (2020)
17. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence **77**(2), 321–358 (1995)
18. Gabbay, D.M., Rodrigues, O.: Equilibrium states in numerical argumentation networks. Logica Universalis **9**(4), 411–473 (2015)
19. Matt, P., Toni, F.: A game-theoretic measure of argument strength for abstract argumentation. In: Proc. of the 11th European Conference on Logics in Artificial Intelligence, (JELIA'08). pp. 285–297 (2008)
20. Pu, F., Luo, J., Zhang, Y., Luo, G.: Argument ranking with categoriser function. In: Proc. of the 7th International Conference on Knowledge Science, Engineering and Management, (KSEM'14). pp. 290–301 (2014)