Post-hoc Explanation of Extension Semantics

Leila Amgoud

Post-hoc Explanation of Extension Semantics

Leila Amgoud

CNRS – IRIT, France
ORCID (Leila Amgoud): https://orcid.org/0000-0002-1838-4271

Abstract. Extension semantics are formal methods that evaluate *acceptability status* of arguments in argumentation graphs where arguments may attack each other. Understanding and explaining their outcomes is of utmost importance in applications like decision making. Consequently, a plethora of works has been devoted to answer questions of the form "why an argument A is (not) accepted under semantics δ". Existing approaches explain the inner working and decision logic of δ. Their explanations refer thus to the semantics's building blocks like attack, defence and admissibility.

This paper complements the existing landscape with a *post-hoc* approach that discloses relationships between argumentation graphs and outputs of a semantics, regardless of its internals. The new approach offers several advantages, namely it explains more acceptability statuses than just the two (accepted, not accepted) considered in the literature, treats all statuses in a similar way, applies to *any* extension semantics that satisfies two key properties (monotonicity and maximality), and provides subsets of attackers, thus the size of its explanations is bounded by the number of attackers of an argument. We characterize the types of attacks that may target an argument. We show that some have no impact on their target's status while others are influential. Then, we introduce three explanation functions that harness influential attacks. One of them provides *sufficient reasons* that guarantee an argument's status while the others identify changes in the graph that guarantee a change of status to any value (*counterfactuals*) or to a specified one (*contrastive*). We show that sufficient reasons are minimal hitting sets of the counterfactuals and vice-versa.

1 Introduction

An argumentation framework consists of a *graph* and a *semantics*. The nodes and edges of the graph are *arguments* (i.e., justifications of claims) and *attack* relations (i.e., conflicts) between arguments respectively. A semantics is a formal method that evaluates the *acceptability status* of every argument in the graph. A sizeable amount of semantics has been proposed in the literature (see [8, 36] for surveys). The very first ones are *extension-based* [22] as they generate sets of arguments, called extensions, that can jointly be acceptable. Some works, like [7, 26], aggregate extensions to assign to every argument an acceptability status taken from the two-valued qualitative scale $\mathbf{T} = \{\text{accepted, non-accepted}\}$. An argument is accepted under a semantics if it belongs to at least one extension, it is non-accepted otherwise. Other papers, including [3, 27, 13], use the richer four-valued scale $\mathbf{S} = \{\text{sceptically accepted, credulously accepted, undecided, rejected}\}$, which refines $\mathbf{T}$. These works distinguish two types of accepted arguments: those that belong to all extensions (sceptical acceptance) from those that belong to some but not all extensions (credulous acceptance). They also distinguish two types of non-accepted arguments: those that are attacked by at least one extension (rejected) from those not (undecided). The most recent semantics are the so-called *gradual* in [13] as they aim to finer-grained evaluations of arguments and thus use even richer scales than $\mathbf{S}$.

Argumentation has been used for solving various theoretical problems [37] including inconsistency handling (eg., [9]), decision making (eg., [4]), persuasion (eg., [10]) and negotiation (eg., [19]). It has also been at the heart of practical problems [6] including argument search engines [39]. Consequently, like any AI model, its explainability received a lot of attention from the argumentation community. Existing works on the topic can roughly be partitioned into three categories. The first category looks for explaining outcomes of AI models that are not argumentation-based. Examples are works that use argumentation techniques to explain machine learning models (eg. [1, 33]). The second category explains outcomes of argumentation-based models like decision systems (eg. [17]) or defeasible logics (eg. [11, 24, 25]). The third category of works investigates explainability of acceptability semantics, namely explaining their evaluations of arguments. Most works of this category focused on extension semantics (eg., [11, 20, 23, 40]) while only a few considered gradual ones [18, 28]. Our work fits within this third category of works with a focus on extension semantics, namely the classical ones from [22].

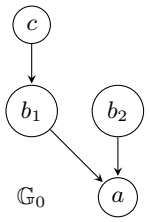
Using the binary scale $\mathbf{T}$, existing works addressed the following questions: why an argument is (not) accepted (eg. [23, 24, 25, 29, 34, 35]) or why a set of arguments is accepted (eg. [38]) under a given semantics. To answer these questions, they followed the so-called *model-based approach* in the XAI literature [31]. Indeed, they explain the inner working and decision logic of the semantics. Consequently, their explanations refer to the semantics building blocks like admissibility and defence. Indeed, an explanation may be an admissible set of arguments [35], a set of arguments defending the targeted argument [11, 29, 40], sub-graphs (or dispute trees) that are sufficient to justify the acceptance status (eg. [23, 24, 35, 38, 20, 21]. These endeavours are similar in spirit to proof procedures (eg., [15, 32]) as they look for simplified ways to check (non-) acceptance of arguments without computing all extensions.

These works are definitely interesting for understanding the behaviour of semantics. On the negative side, they provide the same explanations to arguments that belong to at least one extension. The same holds for arguments that do not belong to any extension. However, we have seen that the scale $\mathbf{S}$ distinguishes two categories of arguments (sceptical, credulous) and (undecided, rejected) in both cases. A precise explanation should be tailored to every acceptability status. Furthermore, while the provided explanations are informative to experts in argumentation, they may not be useful for non-expert

users who are not necessarily interested in learning how semantics behave, but rather look for what led to decisions affecting them. Consider the following multiple criteria decision problem.

Example 1. *Assume a committee in charge of hiring a post-doc researcher to work on an NLP project. Its members exchanged the following arguments concerning Paul's application.*

a**:** *Paul deserves the position since he has a good record.*
b_1**:** *He has no publication in highly selective conferences.*
b_2**:** *He doesn't work on generative AI.*
c**:** *He published in ACL, one of the major NLP conferences.*



The above graph $\mathbb{G}_0$ depicts the attacks between the four arguments. Its grounded extension is $\{b_2, c\}$, thus a is rejected. Assume that Paul didn't get the position. Paul will more likely be interested in knowing the criteria that worked against his application than discovering the grounded extension of the exchanged arguments by the members of the committee. So, a reasonable explanation for him would be the fact that he doesn't work on generative AI (thus the attacker b_2).

This paper complements the existing landscape of models explaining extension semantics by proposing a novel approach which offers several advantages. First, it uses the rich acceptability status scale $\mathbf{S}$ rather than the binary one (accepted, not accepted) considered in the literature. Second, it tailors explanations based on acceptability status. Hence, explanations of a sceptically accepted argument may differ from those of a credulously accepted one, and explanations of an undecided argument may differ from those of a rejected one. Third, it explains in a **unified way** any acceptability status while existing works treat separately the cases of acceptance and non-acceptance. Fourth, it treats any extension semantics that satisfies the *monotonicity* and *maximality* properties, including those from [22], providing thus a **general approach** for explaining a whole family of semantics.

Another **fundamental** difference with existing works lies in the fact that the new approach is *post-hoc* in nature and not model-based [31]. Indeed, it discloses relationships between argumentation graphs and outputs of a semantics, regardless of its internals. Its explanations do not refer to building blocks of semantics, they are rather subsets of argument's direct attackers chosen with great care (eg., b_2 in the above example). It has been shown in [3] that direct attackers are the main culprit of their targets weakening.

The paper starts by presenting a full characterization of all types of attacks that may target an argument. It shows that some are *dummy* in the sense that they have no impact on their target's status while others are *influential*. Among influential attacks, it distinguishes *sufficient attacks*, which guarantee the status of their target under a semantics, and *necessary attacks* whose removal from the graph ensures a status change. Finally, the paper introduces three explanation functions which answer the questions "**why** argument A gets status x under semantics δ?", "**why not** a status other than x for argument A under semantics δ?" and "**why not** the status $\mathtt{s}$ instead of x for argument A under semantics δ?" respectively. The first function returns *sufficient reasons* using sufficient attacks while the second generates *counterfactuals* based on necessary attacks. The two functions seem very different as they answer different questions, however we show that they are closely related. Indeed, sufficient reasons are minimal hitting sets of the counterfactuals and vice-versa. Based on influential attacks, the third function provides *contrastive* explanations, which are changes in the graph that guarantee sceptical acceptance of an argument.

The paper is organized as follows: Section 2 recalls extension semantics and key properties, Section 3 investigates the types of attacks that target an argument, Section 4 introduces three explanation functions, and the last section concludes. The proofs are available in [2].

2 Extension Semantics

The backbone of argumentation is the notion of *argument*, which is a reason behind a standpoint. Arguments may be linked by an *attack relation*, which represents conflicts between pairs of arguments. In the paper, we call *argumentation graph* any pair made of a non-empty finite set of arguments and an attack relation between them. We use the term graph since arguments and their conflicts are represented graphically. Let $\mathtt{Args}$ denote the set of all possible arguments.

Definition 1. *An* argumentation graph *(AG) is a pair $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle$, where $\mathcal{A} \subseteq_f \mathtt{Args}$[1], $\mathcal{A} \neq \emptyset$, and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ (called attack relation). Let $\mathtt{AG}$ be the set of all possible argumentation graphs.*

A pair $(b, a) \in \mathcal{R}$ reads as follows: "b attacks a" or b is an **attacker** of a. We say also that a set $\mathcal{E} \subseteq \mathcal{A}$ attacks $a \in \mathcal{A}$ if $\exists b \in \mathcal{E}$ such that $(b, a) \in \mathcal{R}$.

Notation: Let $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$, $a \in \mathcal{A}$. The function $\mathtt{Att}(a, \mathbb{G})$ returns all attacks targeting a in $\mathbb{G}$ (i.e., $\mathtt{Att}(a, \mathbb{G}) = \{(b, a) \mid (b, a) \in \mathcal{R}\}$). For instance, $\mathtt{Att}(a, \mathbb{G}_0) = \{(b_1, a), (b_2, a)\}$.

An *extension semantics* is a formal method that evaluates the *acceptability status* of arguments in argumentation graphs. It looks for sets of arguments, called *extensions*, that are jointly acceptable.

Definition 2. *An* extension semantics *is a function δ mapping every $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$ into $\mathtt{Ext}_\delta(\mathbb{G}) \subseteq \mathbb{P}(\mathcal{A})$[2]. Every member of $\mathtt{Ext}_\delta(\mathbb{G})$ is called* extension.

Examples of extension semantics are those defined in [22], namely complete (*co*), grounded (*gr*), preferred (*pr*), and stable (*st*). They are based on the notions of *conflict-freeness* and *defence*, where for an argumentation graph $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle$, $\mathcal{E} \subseteq \mathcal{A}$, and an argument $a \in \mathcal{A}$,

- $\mathcal{E}$ is *conflict-free* iff $\nexists a, b \in \mathcal{E}$ such that $(a, b) \in \mathcal{R}$.
- $\mathcal{E}$ *defends* a iff $\forall b \in \mathcal{A}$ such that $(b, a) \in \mathcal{R}$, $\exists c \in \mathcal{E}$ such that $(c, b) \in \mathcal{R}$.

Let us now recall the definition of an extension under the abovementioned four semantics.

- $\mathcal{E}$ is a *complete* extension iff $\mathcal{E}$ is conflict-free, defends its elements and contains all the arguments it defends.
- $\mathcal{E}$ is a *grounded* extension iff it is the subset-minimal complete extension.
- $\mathcal{E}$ is a *preferred* extension iff it is a subset-maximal complete extension.
- $\mathcal{E}$ is a *stable* extension iff $\mathcal{E}$ is conflict-free and for any argument $a \in \mathcal{A} \setminus \mathcal{E}$, $\mathcal{E}$ attacks a.

Example 2. *Let us analyse the six graphs from Figure 1 under the preferred semantics.*

[1] The notation $\mathcal{A} \subseteq_f \mathtt{Args}$ means $\mathcal{A}$ is a finite subset of $\mathtt{Args}$.
[2] $\mathbb{P}(\mathcal{A})$ is the *powerset* - the set of all partitions - of $\mathcal{A}$.

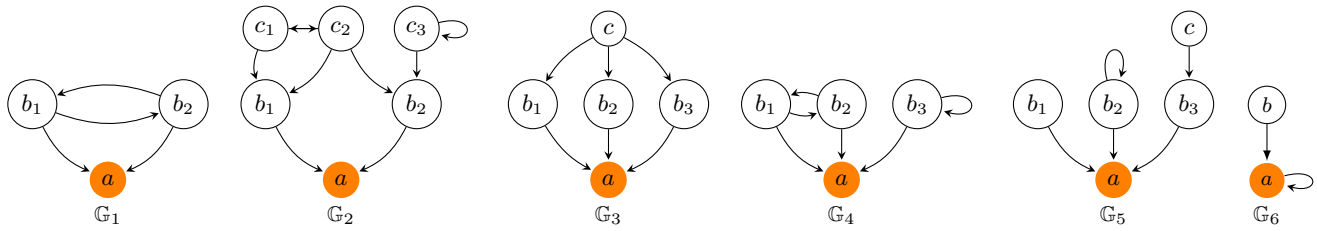


Figure 1. Examples of argumentation graphs $\mathbb{G}_i = \langle \mathcal{A}_i, \mathcal{R}_i \rangle, i \in \{1, \ldots, 6\}$.

- $\texttt{Ext}_{pr}(\mathbb{G}_1) = \{\{b_1\}, \{b_2\}\}$.
- $\texttt{Ext}_{pr}(\mathbb{G}_2) = \{\{c_1\}, \{a, c_2\}\}$.
- $\texttt{Ext}_{pr}(\mathbb{G}_3) = \{\{a, c\}\}$.
- $\texttt{Ext}_{pr}(\mathbb{G}_4) = \{\{b_1\}, \{b_2\}\}$.
- $\texttt{Ext}_{pr}(\mathbb{G}_5) = \{\{b_1, c\}\}$.
- $\texttt{Ext}_{pr}(\mathbb{G}_6) = \{\{b\}\}$.

Extension semantics may return 0, 1, or more extensions. The latter are aggregated in order to assign an *acceptability status* to every individual argument in the analysed graph. In this paper, we will use the four-valued scale $\mathbf{S} = \{\texttt{s}, \texttt{c}, \texttt{u}, \texttt{r}\}$ for the purpose. An argument is sceptically accepted (s) if it belongs to all extensions, it is credulously accepted (c) if it belongs to some but not all extensions. When an argument does not belong to any extension, it is said to be undecided (u) if it is not attacked by any extension and it is rejected (r) if it is attacked by at least one extension. A question that raises naturally is: what is the status of an argument when a semantics (like stable) fails to return extensions? Consider the graph $\mathbb{G}_2$, which has no stable extension and let us focus on the status of a. Some works like [23, 35] consider as non-accepted any argument that does not belong to any extension. Hence, a is non-accepted in a *vacuous way*. However, this is not a fair evaluation since it is not based on evidence, i.e., concrete extensions. In fact, when a semantics fails to return extensions, no evaluation is performed and thus arguments keep intact their strength. In this paper, we consider that every argument is presumed strong (s) until proven otherwise.

Throughout the paper, the scale $\mathbf{S}$ is equipped with a total order $\succeq$ such that $\texttt{s} \succ \texttt{c} \succ \texttt{u} \succ \texttt{r}$, where $x \succeq y$ and $x \succ y$ mean "x is as acceptable as y" and "x is more acceptable than y" respectively.

Definition 3. *An* aggregator *based on extension semantics δ is a mapping $\texttt{Acc}_\delta$ from $\texttt{Args} \times \texttt{AG}$ to $\mathbf{S}$ such that for any $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, for any $a \in \mathcal{A}$, if $\texttt{Ext}_\delta(\mathbb{G}) = \emptyset$, then $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{s}$. Else:*

- $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{s}$ *iff* $a \in \bigcap_{\mathcal{E} \in \texttt{Ext}_\delta(\mathbb{G})} \mathcal{E}$.
- $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{c}$ *iff* $\exists \mathcal{E}, \mathcal{E}' \in \texttt{Ext}_\delta(\mathbb{G})$ *such that* $a \in \mathcal{E}$, $a \notin \mathcal{E}'$.
- $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{r}$ *iff* $a \notin \bigcup_{\mathcal{E} \in \texttt{Ext}_\delta(\mathbb{G})} \mathcal{E}$ *and* $\exists \mathcal{E} \in \texttt{Ext}_\delta(\mathbb{G})$ *such that* $\mathcal{E}$ *attacks* a.
- $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{u}$ *iff* $a \notin \bigcup_{\mathcal{E} \in \texttt{Ext}_\delta(\mathbb{G})} \mathcal{E}$ *and* $\nexists \mathcal{E} \in \texttt{Ext}_\delta(\mathbb{G})$ *such that* $\mathcal{E}$ *attacks* a.

$\texttt{Acc}_\delta(a, \mathbb{G})$ *denotes the* acceptability status *of the argument a in the graph $\mathbb{G}$ under the semantics δ.*

Example 2 (Cont). Let us analyse the status of every argument of the six graphs from Figure 1 under preferred semantics.

- $\texttt{Acc}_{pr}(b_1, \mathbb{G}_1) = \texttt{Acc}_{pr}(b_2, \mathbb{G}_1) = \texttt{c}$, and $\texttt{Acc}_{pr}(a, \mathbb{G}_1) = \texttt{r}$.
- $\texttt{Acc}_{pr}(a, \mathbb{G}_2) = \texttt{Acc}_{pr}(c_1, \mathbb{G}_2) = \texttt{Acc}_{pr}(c_2, \mathbb{G}_2) = \texttt{c}$, $\texttt{Acc}_{pr}(c_3, \mathbb{G}_2) = \texttt{u}$, $\texttt{Acc}_{pr}(b_1, \mathbb{G}_2) = \texttt{Acc}_{pr}(b_2, \mathbb{G}_2) = \texttt{r}$.
- $\texttt{Acc}_{pr}(a, \mathbb{G}_3) = \texttt{Acc}_{pr}(c, \mathbb{G}_3) = \texttt{s}$ and for any $i \in \{1, 2, 3\}$, $\texttt{Acc}_{pr}(b_i, \mathbb{G}_3) = \texttt{r}$.
- $\texttt{Acc}_{pr}(b_1, \mathbb{G}_4) = \texttt{Acc}_{pr}(b_2, \mathbb{G}_4) = \texttt{c}$, $\texttt{Acc}_{pr}(b_3, \mathbb{G}_4) = \texttt{u}$, and $\texttt{Acc}_{pr}(a, \mathbb{G}_4) = \texttt{r}$.
- $\texttt{Acc}_{pr}(b_1, \mathbb{G}_5) = \texttt{Acc}_{pr}(c, \mathbb{G}_5) = \texttt{s}$, $\texttt{Acc}_{pr}(b_2, \mathbb{G}_5) = \texttt{u}$, $\texttt{Acc}_{pr}(a, \mathbb{G}_5) = \texttt{Acc}_{pr}(b_3, \mathbb{G}_5) = \texttt{r}$.
- $\texttt{Acc}_{pr}(a, \mathbb{G}_6) = \texttt{r}$, $\texttt{Acc}_{pr}(b, \mathbb{G}_6) = \texttt{s}$.

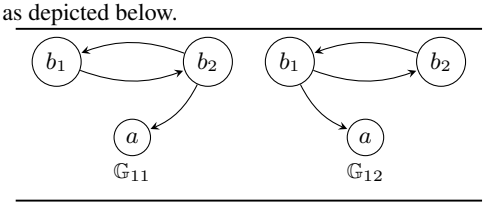
Let us now introduce two functions of sub-graphs, which are useful for the rest of our study. They consist both of removing some attacks from the graph while keeping the whole set of arguments.

Definition 4. *Let $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$, and $X \subseteq \texttt{Att}(a, \mathbb{G})$. We define $\mathbb{G} \ominus X$ and $\mathbb{G} \downarrow X$ as two argumentation graphs such that:*

- $\mathbb{G} \ominus X = \langle \mathcal{A}, \mathcal{R} \setminus X \rangle$,
- $\mathbb{G} \downarrow X = \langle \mathcal{A}, \mathcal{R}' \rangle$ *where* $\texttt{Att}(a, \mathbb{G} \downarrow X) = X$ *and* $\forall b \in \mathcal{A} \setminus \{a\}$, $\texttt{Att}(b, \mathbb{G} \downarrow X) = \texttt{Att}(b, \mathbb{G})$.

Hence, the graph $\mathbb{G} \ominus X$ is $\mathbb{G}$ where all attacks in X are removed and the graph $\mathbb{G} \downarrow X$ is $\mathbb{G}$ but where the set of attacks on the argument a is limited to those in X.

Example 2 (Cont). Consider the graph $\mathbb{G}_1$ in Figure 1 and the argument a. Note that $\texttt{Att}(a, \mathbb{G}_1) = \{(b_1, a), (b_2, a)\}$. For $X = \{(b_1, a)\}$, $\mathbb{G}_1 \ominus X = \mathbb{G}_{11}$ and $\mathbb{G}_1 \downarrow X = \mathbb{G}_{12}$, where $\mathbb{G}_{11}$ and $\mathbb{G}_{12}$ are as depicted below.



Let us now recall two formal properties from the literature that a semantics should satisfy. Introduced in [3], the *Maximality* property states that any non-attacked argument does not loose strength. This means that it is presumed sceptically accepted.

Definition 5 (Maximality)**.** *An extension semantics δ satisfies* maximality *iff for any $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, for any $a \in \mathcal{A}$, if $\texttt{Att}(a, \mathbb{G}) = \emptyset$, then $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{s}$.*

Example 1 (Cont). Any extension semantics that satisfies Maximality assigns s to b_2 and c in the graph $\mathbb{G}_0$.

The second property, called *Monotonicity*, has been defined in [5] and states that removing attacks to an argument cannot lower the acceptability status of the argument. Indeed, the status can either remain unchanged or increase in the modified graph.

Definition 6 (Monotonicity)**.** *A semantics δ is* monotonic *iff for any $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, for any $a \in \mathcal{A}$, for any $X \subseteq \texttt{Att}(a, \mathbb{G})$,*

$$\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) \succeq \texttt{Acc}_\delta(a, \mathbb{G}).$$

Example 2 (Cont). Consider the graph $\mathbb{G}_1$ in Figure 1 and the argument a. For any $X \in \{\emptyset, \{(b_1,a)\}, \{(b_2,a)\}, \{(b_1,a), (b_2,a)\}\}$, for any extension semantics δ which is monotonic, it holds that $\texttt{Acc}_\delta(a, \mathbb{G}_1 \ominus X) \succeq \texttt{Acc}_\delta(a, \mathbb{G}_1)$.

It has been shown that all the reviewed semantics (complete, grounded, stable, preferred) are monotonic [5] and satisfy Maximality [3] (following Def. 3 for status assignment).

Theorem 1. *[3, 5] For any $\delta \in \{co, gr, st, pr\}$, δ satisfies Maximality and Monotonicity.*

From the monotonicity property of extension semantics follows a monotonic behaviour of the acceptability status of arguments. Indeed, we show that if the status of an argument in a graph is the same in a subgraph where only a subset of its attacks is considered, then the same status is guaranteed in any subgraph considering a superset of attacks. This shows that a subset of attacks is sufficient to guarantee the current status. In a similar way, we show that if removing some attacks will lead to a change in the status of the argument, then removing any superset of attacks would also lead to a change.

Theorem 2. *Let δ be a monotonic semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$ and $X \subseteq \texttt{Att}(a, \mathbb{G})$.*

1. *If $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X)$, then $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X')$ for any $X \subset X' \subseteq \texttt{Att}(a, \mathbb{G})$.*
2. *If $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus X)$, then $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus X')$ for any $X \subset X' \subseteq \texttt{Att}(a, \mathbb{G})$.*

To sum up, Monotonicity and Maximality show that direct attacks on an argument are responsible for its loss of strength. Hence, to explain the status of any argument, one should focus on its attackers.

3 Characterization of Types of Attacks

Let us characterize the different *types* of attacks that can target an argument, investigate their links, and clarify their roles under any extension semantics that satisfies Monotonicity and Maximality. We show that an attack is either *dummy* or *influential*. In the first case, it has no impact on its target's acceptability status while in the second it does. Influential attacks are of three categories: *necessary*, *sufficient*, and *inhibited*. The later are harmful only when they are not accompanied with necessary/sufficient attacks.

3.1 Dummy Attacks

A dummy attack has no impact on its target's acceptability status as its removal from the graph does not result in a status change. However, this is not sufficient to guarantee that the attack is dummy. Consider the argumentation graph $\mathbb{G}_5$ and let us focus on the attack $r = (b_2, a)$. Recall that a is rejected under preferred semantics (i.e., $\texttt{Acc}_{pr}(a, \mathbb{G}_5) = \texttt{r}$). Note that if r is removed from the graph, a remains rejected, i.e., $\texttt{Acc}_{pr}(a, \mathbb{G}_5 \ominus \{(b_2,a)\}) = \texttt{r}$. Hence, one may think that r is dummy. This is not the case since it may have impact on a in the subgraph $\mathbb{G}_5 \ominus \{(b_1,a), (b_3,a)\})$ since $\texttt{Acc}_{pr}(a, \mathbb{G}_5 \ominus \{(b_1,a), (b_3,a)\}) = \texttt{u}$. Hence, to check whether an attack is dummy, one should verify its **marginal contribution** alone and in presence of any other subset of attacks.

Definition 7 (Dummy)**.** *Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$, and $r \in \texttt{Att}(a, \mathbb{G})$. The attack r is* dummy *for $\texttt{Acc}_\delta(a, \mathbb{G})$ iff $\forall X \subseteq \texttt{Att}(a, \mathbb{G}) \setminus \{r\}$,*

$$\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus (X \cup \{r\})).$$

Let $\texttt{Dum}_\delta(a, \mathbb{G})$ be the set of dummy attacks on a in $\mathbb{G}$ under δ.

Example 2 (Cont). Let us check the dummy attacks of the argument a in each graph $\mathbb{G}_i$ in Figure 1 under preferred semantics.

- $\texttt{Dum}_{pr}(a, \mathbb{G}_1) = \emptyset$
- $\texttt{Dum}_{pr}(a, \mathbb{G}_2) = \{(b_1, a)\}$
- $\texttt{Dum}_{pr}(a, \mathbb{G}_3) = \texttt{Att}(a, \mathbb{G}_3)$
- $\texttt{Dum}_{pr}(a, \mathbb{G}_4) = \emptyset$
- $\texttt{Dum}_{pr}(a, \mathbb{G}_5) = \{(b_3, a)\}$
- $\texttt{Dum}_{pr}(a, \mathbb{G}_6) = \emptyset$

Note that the three attacks $r_i = (b_i, a)$ in the graph $\mathbb{G}_3$ are all dummy. Consider for instance the attack r_3 and let $X_1 = \emptyset$, $X_2 = \{r_1\}$, $X_3 = \{r_2\}$ and $X_4 = \{r_1, r_2\}$. Note that:

- $\texttt{Acc}_{pr}(a, \mathbb{G}_3 \ominus X_1) = \texttt{Acc}_{pr}(a, \mathbb{G}_3 \ominus (X_1 \cup \{r_3\})) = \texttt{s}$,
- $\texttt{Acc}_{pr}(a, \mathbb{G}_3 \ominus X_2) = \texttt{Acc}_{pr}(a, \mathbb{G}_3 \ominus (X_2 \cup \{r_3\})) = \texttt{s}$,
- $\texttt{Acc}_{pr}(a, \mathbb{G}_3 \ominus X_3) = \texttt{Acc}_{pr}(a, \mathbb{G}_3 \ominus (X_3 \cup \{r_3\})) = \texttt{s}$,
- $\texttt{Acc}_{pr}(a, \mathbb{G}_3 \ominus X_4) = \texttt{Acc}_{pr}(a, \mathbb{G}_3 \ominus (X_4 \cup \{r_3\})) = \texttt{s}$.

Remark. Note that an attack whose source is rejected is not necessarily dummy. For instance, the argument a of the graph $\mathbb{G}_6$ is rejected while the self-attack (a, a) is not dummy.

We show next that since dummy attacks have no impact on the status of their targets, they can safely be removed from the graph. Note that the result holds for any semantics, be it monotonic or not.

Proposition 3. *Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$. For any $a \in \mathcal{A}$, for any $X \subseteq \texttt{Dum}_\delta(a, \mathbb{G})$, $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X)$.*

3.2 Influential Attacks

Influential attacks are those that, if removed, may result in a change in the status of their targets. They have impact on acceptability status, be them alone or in presence of some other subset of attacks. These are therefore attacks which are not dummy.

Definition 8 (Influential)**.** *Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$, and $r \in \texttt{Att}(a, \mathbb{G})$. The attack r is* influential *for $\texttt{Acc}_\delta(a, \mathbb{G})$ iff $\exists X \subseteq \texttt{Att}(a, \mathbb{G}) \setminus \{r\}$ such that:*

$$\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus (X \cup \{r\})).$$

Let $\texttt{Inf}_\delta(a, \mathbb{G})$ be the set of influential attacks of a in $\mathbb{G}$ under δ.

Example 2 (Cont). Let us check the influential attacks of the argument a in each graph in the Figure 1 under preferred semantics.

- $\texttt{Inf}_{pr}(a, \mathbb{G}_1) = \texttt{Att}(a, \mathbb{G}_1)$
- $\texttt{Inf}_{pr}(a, \mathbb{G}_2) = \{(b_2, a)\}$
- $\texttt{Inf}_{pr}(a, \mathbb{G}_3) = \emptyset$
- $\texttt{Inf}_{pr}(a, \mathbb{G}_4) = \texttt{Att}(a, \mathbb{G}_4)$
- $\texttt{Inf}_{pr}(a, \mathbb{G}_5) = \{(b_1, a), (b_2, a)\}$
- $\texttt{Inf}_{pr}(a, \mathbb{G}_6) = \texttt{Att}(a, \mathbb{G}_6)$

Note that the argument a in the graph $\mathbb{G}_6$ is rejected (under preferred semantics). This rejection is due to the attack coming from b. Thus, one may think that the self-attack of a does not have any role as its removal would not change the status of a. In fact, the self-attack has influence on a since if the attack from b is removed, then a would be undecided while removing both makes a sceptically accepted.

From the above definitions, it follows that the attacks of an argument are either dummy or influential.

Proposition 4. *Let δ be a semantics and $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$. For any $a \in \mathcal{A}$, $\mathtt{Inf}_\delta(a, \mathbb{G}) = \mathtt{Att}(a, \mathbb{G}) \setminus \mathtt{Dum}_\delta(a, \mathbb{G})$.*

We show that influential attacks prevent their targets from being sceptically accepted under any semantics which satisfies Maximality and Monotonicity, including thus the four reviewed semantics.

Theorem 5. *Let δ be an extension semantics which satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$ and $a \in \mathcal{A}$ such that $\mathtt{Inf}_\delta(a, \mathbb{G}) \neq \emptyset$.*

- $\mathtt{Acc}_\delta(a, \mathbb{G}) \neq \mathtt{s}$,
- $\mathtt{Acc}_\delta(a, \mathbb{G} \ominus \mathtt{Inf}_\delta(a, \mathbb{G})) = \mathtt{s}$.

In the next sub-sections, we discuss the three categories of influential attacks and their respective roles.

3.2.1 Necessary Attacks

Among influential attacks of an argument, we distinguish *necessary* attacks whose removal from the argumentation graph results automatically in a change in the argument's status. Put differently, their absence would lead to another status for the argument. In what follows, we define subset-minimal necessary attacks.

Definition 9 (Necessity)**.** *Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$, $a \in \mathcal{A}$ and $X \subseteq \mathtt{Att}(a, \mathbb{G})$. The set X is* necessary *for $\mathtt{Acc}_\delta(a, \mathbb{G})$ iff:*

- $\mathtt{Acc}_\delta(a, \mathbb{G}) \neq \mathtt{Acc}_\delta(a, \mathbb{G} \ominus X)$,
- $\nexists X' \subset X$ *such that X' satisfies the above condition.*

$\mathtt{Nec}_\delta(a, \mathbb{G})$ *is the set of all necessary sets for $\mathtt{Acc}_\delta(a, \mathbb{G})$ under δ.*

Example 2 (Cont). Let us check the necessary attacks for the status of the argument a in each graph under preferred semantics.

- $\mathtt{Nec}_{pr}(a, \mathbb{G}_1) = \{\{(b_1, a)\}, \{(b_2, a)\}\}$
- $\mathtt{Nec}_{pr}(a, \mathbb{G}_2) = \{\{(b_2, a)\}\}$
- $\mathtt{Nec}_{pr}(a, \mathbb{G}_3) = \emptyset$
- $\mathtt{Nec}_{pr}(a, \mathbb{G}_4) = \{\{(b_1, a), (b_2, a)\}, \{(b_1, a), (b_3, a)\}, \{(b_2, a), (b_3, a)\}\}$
- $\mathtt{Nec}_{pr}(a, \mathbb{G}_5) = \{\{(b_1, a)\}\}$
- $\mathtt{Nec}_{pr}(a, \mathbb{G}_6) = \{\{(b, a)\}\}$

For instance, the argument a in $\mathbb{G}_1$ has two necessary sets of attacks, each of which contains one of the two attacks on a. Note also that necessary attacks **may not exist**. Indeed, $\mathtt{Nec}_{pr}(a, \mathbb{G}_3) = \emptyset$ meaning that the three attacks can safely be removed without altering the status of a. Finally, note that in the graph $\mathbb{G}_2$, the attack (b_2, a) is necessary for the argument a while it emanates from a rejected argument ($\mathtt{Acc}_{pr}(b_2, \mathbb{G}_2) = \mathtt{r}$). Hence, a necessary attack does not necessarily come from a strong argument.

We show that if the removal of a subset of an argument's attacks results in a change in the argument's status, then the removed set contains a necessary set. This result holds for any extension semantics.

Proposition 6. *Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$, $a \in \mathcal{A}$ such that $\mathtt{Att}(a, \mathbb{G}) \neq \emptyset$. For any $X \subseteq \mathtt{Att}(a, \mathbb{G})$, if $\mathtt{Acc}_\delta(a, \mathbb{G}) \neq \mathtt{Acc}_\delta(a, \mathbb{G} \ominus X)$, then $\exists X' \subseteq X$ such that $X' \in \mathtt{Nec}_\delta(a, \mathbb{G})$.*

The running example, namely through the graph $\mathbb{G}_3$, shows that necessary sets may not exist ($\mathtt{Nec}_{pr}(a, \mathbb{G}_3) = \emptyset$). The following result states that when they do exist, they cannot be empty.

Proposition 7. *Let δ be an extension semantics and $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$. For any $a \in \mathcal{A}$, $\emptyset \notin \mathtt{Nec}_\delta(a, \mathbb{G})$.*

3.2.2 Sufficient Attacks

The second category of influential attacks is that of *sufficient attacks*. A set of attacks on an argument is sufficient for the argument's status if considered alone (removing all the remaining attacks towards the argument) leads to the same status.

Definition 10 (Sufficiency)**.** *Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$, $a \in \mathcal{A}$, and $X \subseteq \mathtt{Att}(a, \mathbb{G})$. The set X is* sufficient *for $\mathtt{Acc}_\delta(a, \mathbb{G})$ iff:*

- $\mathtt{Acc}_\delta(a, \mathbb{G}) = \mathtt{Acc}_\delta(a, \mathbb{G} \downarrow X)$.
- $\nexists X' \subset X$ *such that X' satisfies the above condition.*

$\mathtt{Suff}_\delta(a, \mathbb{G})$ *is the set of all sufficient sets for $\mathtt{Acc}_\delta(a, \mathbb{G})$ under δ.*

Example 2 (Cont). Let us check the sufficient attacks for the status of the argument a in each graph under preferred semantics.

- $\mathtt{Suff}_{pr}(a, \mathbb{G}_1) = \{\{(b_1, a), (b_2, a)\}\}$
- $\mathtt{Suff}_{pr}(a, \mathbb{G}_2) = \{\{(b_2, a)\}\}$
- $\mathtt{Suff}_{pr}(a, \mathbb{G}_3) = \{\emptyset\}$
- $\mathtt{Suff}_{pr}(a, \mathbb{G}_4) = \{\{(b_1, a), (b_2, a)\}, \{(b_1, a), (b_3, a)\}, \{(b_2, a), (b_3, a)\}\}$
- $\mathtt{Suff}_{pr}(a, \mathbb{G}_5) = \{\{(b_1, a)\}\}$
- $\mathtt{Suff}_{pr}(a, \mathbb{G}_6) = \{\{(b, a)\}\}$

Note that the argument a has a single sufficient set in $\mathbb{G}_1$ while it has two necessary sets. The idea is that to be rejected, both attacks on a should be present in the graph. The removal of each of them leads to another status for a.

A first property states that if the set of attacks on an argument is reduced while keeping the argument's status, then the retained attacks contain necessarily a sufficient set.

Proposition 8. *Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$, and $a \in \mathcal{A}$. For any $X \subseteq \mathtt{Att}(a, \mathbb{G})$, if $\mathtt{Acc}_\delta(a, \mathbb{G}) = \mathtt{Acc}_\delta(a, \mathbb{G} \downarrow X)$, then $\exists X' \subseteq X$ such that $X' \in \mathtt{Suff}_\delta(a, \mathbb{G})$.*

We show next that unlike necessary attacks, sufficient sets do exist.

Proposition 9. *Let δ be an extension semantics and $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$. For any $a \in \mathcal{A}$, $\mathtt{Suff}_\delta(a, \mathbb{G}) \neq \emptyset$.*

Again unlike necessary sets which cannot be empty, the empty set can be sufficient for the status of an argument. This occurs when the latter is sceptically accepted and all its attacks are dummy under any extension semantics that satisfies Maximality and Monotonicity.

Theorem 10. *Let δ be an extension semantics which satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \mathtt{AG}$, and $a \in \mathcal{A}$. The statements below are pairwise equivalent.*

1. $\mathtt{Acc}_\delta(a, \mathbb{G}) = \mathtt{s}$,
2. $\mathtt{Suff}_\delta(a, \mathbb{G}) = \{\emptyset\}$,
3. $\mathtt{Dum}_\delta(a, \mathbb{G}) = \mathtt{Att}(a, \mathbb{G})$.

Example 2 (Cont). Consider the graph $\mathbb{G}_3$ from Figure 1. Recall that $\texttt{Acc}_{pr}(a, \mathbb{G}_3) = \texttt{s}$. It can be checked that $\forall X \subset \texttt{Att}(a, \mathbb{G}_3) = \{(b_i, a) \mid i = 1, 2, 3\}$, $\texttt{Acc}_{pr}(a, \mathbb{G}_3 \ominus X) = \texttt{s}$. So, $\texttt{Suff}_\delta(a, \mathbb{G}_3) = \{\emptyset\}$ and $\texttt{Dum}_\delta(a, \mathbb{G}_3) = \texttt{Att}(a, \mathbb{G}_3)$.

The following result establishes a first link between sufficient sets and necessary ones for the status of an argument under an extension semantics that satisfies Maximality and Monotonicity. It shows that non-existence of necessary sets occurs only when the empty set is the only sufficient set.

Theorem 11. *Let δ be an extension semantics that satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$ and $a \in \mathcal{A}$.*

$$\texttt{Suff}_\delta(a, \mathbb{G}) = \{\emptyset\} \textit{ iff } \texttt{Nec}_\delta(a, \mathbb{G}) = \emptyset.$$

From the above result, it follows that the empty set is the only sufficient set for any non-attacked argument.

Corollary 12. *Let δ be an extension semantics that satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$ and $a \in \mathcal{A}$. If $\texttt{Att}(a, \mathbb{G}) = \emptyset$, then $\texttt{Suff}_\delta(a, \mathbb{G}) = \{\emptyset\}$ and $\texttt{Nec}_\delta(a, \mathbb{G}) = \emptyset$.*

Despite their different roles, we show next that the notions of necessity and sufficiency are **dual**. Indeed, a sufficient set for the status of an argument is a minimal **hitting set** of the necessary sets for the status of the argument and vice-versa. Put differently, a necessary set for the status of an argument is a minimal subset of its attacks that contains at least one element from every sufficient set. A sufficient set is a minimal subset of attacks which contains at least one element from every necessary set. Hence, each notion can be generated from the other. This result holds for any extension semantics that satisfies the two properties of Maximality and Monotonicity.

Theorem 13. *Let δ be an extension semantics that satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$ and $a \in \mathcal{A}$ such that $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$.*

- *$X \in \texttt{Suff}_\delta(a, \mathbb{G})$ iff X is a subset-minimal of $\texttt{Att}(a, \mathbb{G})$ such that $\forall Y \in \texttt{Nec}_\delta(a, \mathbb{G})$, $X \cap Y \neq \emptyset$.*
- *$X \in \texttt{Nec}_\delta(a, \mathbb{G})$ iff X is a subset-minimal of $\texttt{Att}(a, \mathbb{G})$ such that $\forall Y \in \texttt{Suff}_\delta(a, \mathbb{G})$, $X \cap Y \neq \emptyset$.*

3.2.3 Inhibited Attacks

An **inhibited** attack is influential but it is neither necessary nor sufficient for its target's status. It is said to be inhibited since it has impact *only* when it is not accompanied with necessary/sufficient attacks.

Definition 11 (Inhibited)**.** *Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$, $r \in \texttt{Att}(a, \mathbb{G})$. The attack r is* inhibited *for $\texttt{Acc}_\delta(a, \mathbb{G})$ iff $r \in \texttt{Inf}_\delta(a, \mathbb{G})$ and $\nexists X \in \texttt{Suff}(a, \mathbb{G})$ such that $r \in X$. Let $\texttt{Inh}_\delta(a, \mathbb{G})$ be the set of all inhibited attacks for $\texttt{Acc}_\delta(a, \mathbb{G})$.*

Example 2 (Cont). In $\mathbb{G}_5$, the attack (b_2, a) is influential for $\texttt{Acc}_{pr}(a, \mathbb{G}_5)$ but it is neither sufficient nor necessary. However, a would be undecided (and not rejected) in $\mathbb{G}_5 \ominus \{(b_1, a)\}$. Thus, $(b_2, a) \in \texttt{Inh}_{pr}(a, \mathbb{G}_5)$. Note also that $(a, a) \in \texttt{Inh}_{pr}(a, \mathbb{G}_6)$.

The following result summarises the links between the different types of attacks. It shows in particular that the set of attacks that belong to sufficient sets coincides with the set containing attacks appearing in necessary sets. Influential set contains two disjoints subsets: inhibited attacks and those that are in sufficient/necessary sets.

Theorem 14. *Let δ be an extension semantics that satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, and $a \in \mathcal{A}$.*

1. $\texttt{Att}(a, \mathbb{G}) = \texttt{Inf}_\delta(a, \mathbb{G}) \cup \texttt{Dum}_\delta(a, \mathbb{G})$, *with* $\texttt{Inf}_\delta(a, \mathbb{G}) \cap \texttt{Dum}_\delta(a, \mathbb{G}) = \emptyset$ *(Proposition 4)*
2. $\Big(\bigcup_{X \in \texttt{Suff}_\delta(a, \mathbb{G})} X\Big) = \Big(\bigcup_{Y \in \texttt{Nec}_\delta(a, \mathbb{G})} Y\Big) \subseteq \texttt{Inf}_\delta(a, \mathbb{G})$,
3. $\texttt{Inh}_\delta(a, \mathbb{G}) = \texttt{Inf}_\delta(a, \mathbb{G}) \setminus \Big(\bigcup_{X \in \texttt{Suff}_\delta(a, \mathbb{G})} X\Big)$.

4 Post-hoc Explanation Functions

Let us now introduce the novel approach for explaining outcomes of extension semantics. Unlike exiting works that describe the internal reasoning of semantics, it looks rather for correlations between inputs (argumentation graphs) and outcomes of a semantics (acceptability status of arguments). We introduce three explanation functions, which answer different questions. Before that, let us first define the notion of query, which is a tuple made of an extension semantics, an argumentation graph and an argument. The idea is to explain the status of the argument in the graph and under the given semantics.

Definition 12 (Query)**.** *A* query *is a tuple $\mathbf{Q} = \langle \delta, \mathbb{G}, a \rangle$ where δ is an extension semantics which satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$ and $a \in \mathcal{A}$.*

An explanation function takes as input a query and returns a set of explanations. In the approach, an explanation is a set of arguments.

Definition 13 (Explainer)**.** *An* explainer *is a function* $\texttt{g}$ *mapping every query $\mathbf{Q} = \langle \delta, \mathbb{G}, a \rangle$ into $\mathbb{P}(\texttt{Args})$.*

Our first explanation function answers questions of the form:

> **[Q1:]** ***Why*** *argument a from graph $\mathbb{G}$ gets status x under semantics δ? Put differently, why $\texttt{Acc}_\delta(a, \mathbb{G}) = x$?*

The answer consists in highlighting key factors that caused the status x. Our explainer identifies the direct attackers of a that are sufficient for ensuring the current status x. An explanation E is therefore a subset of attackers of a whose attacks constitute a sufficient set. It reads as: $\texttt{Acc}_\delta(a, \mathbb{G}) = x$ *because of the arguments in* E.

Definition 14 ($\texttt{g}_f$)**.** *A* factual explainer *is a function $\texttt{g}_f$ such that for every query $\mathbf{Q} = \langle \delta, \mathbb{G}, a \rangle$ with $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle$,*

$$\texttt{g}_f(\mathbf{Q}) = \{E \subseteq \texttt{Args} \mid \{(b_i, a) | \, b_i \in E\} \in \texttt{Suff}_\delta(a, \mathbb{G})\}.$$

Example 1 (Cont). The graph $\mathbb{G}_0$ has one preferred extension: $\{b_2, c\}$. Thus, $\texttt{Acc}_{pr}(a, \mathbb{G}_0) = \texttt{r}$, $\texttt{Suff}_{pr}(a, \mathbb{G}_0) = \texttt{Nec}_{pr}(a, \mathbb{G}_0) = \{\{(b_2, a)\}\}$. It follows that for $\mathbf{Q}_0 = \langle \text{pr}, \mathbb{G}_0, a \rangle$, $\texttt{g}_f(\mathbf{Q}_0) = \{\{b_2\}\}$, i.e., the reason for rejecting Paul's application is the fact that Paul doesn't work on generative AI.

Example 2 (Cont). Let $\mathbf{Q}_i = \langle \text{pr}, \mathbb{G}_i, a \rangle$ with $i \in \{1, \ldots, 6\}$.

- $\texttt{g}_f(\mathbf{Q}_1) = \{\{b_1, b_2\}\}$
- $\texttt{g}_f(\mathbf{Q}_2) = \{\{b_2\}\}$
- $\texttt{g}_f(\mathbf{Q}_3) = \{\emptyset\}$
- $\texttt{g}_f(\mathbf{Q}_4) = \{\{b_1, b_2\}, \{b_1, b_3\}, \{b_2, b_3\}\}$
- $\texttt{g}_f(\mathbf{Q}_5) = \{\{b_1\}\}$
- $\texttt{g}_f(\mathbf{Q}_6) = \{\{b\}\}$

The second explanation function answers the following question:

[**Q2:**] ***Why not*** *a status other than x for argument a from graph $\mathbb{G}$ under semantics δ? i.e., why* $\texttt{Acc}_\delta(a, \mathbb{G}) \notin \mathbf{S} \setminus \{x\}$*?*

Note that this question **does not specify a desirable value** from $\mathbf{S} \setminus \{x\}$ for the argument a. To answer such questions, our second function generates *counterfactual* explanations that state how the graph $\mathbb{G}$ would have to be different to get a status other than x. In other words, they identify the changes that should minimally occur for altering the status of a to **any other value** in the scale $\mathbf{S}$. Our function focuses on necessary attacks as we have seen previously that their removal from the graph results in a status change. Every explanation E reads as follows: *If the attacks in E had not been present in $\mathbb{G}$, the status of a would have been different from x.*

Definition 15 (g_c)**.** *A* counterfactual explainer *is a function* g_c *such that for every query* $\mathbf{Q} = \langle \delta, \mathbb{G}, a \rangle$*, with* $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle$*,*

$$\mathsf{g}_c(\mathbf{Q}) = \{E \subseteq \texttt{Args} \mid \{(b_i, a) | \, b_i \in E\} \in \texttt{Nec}_\delta(a, \mathbb{G})\}.$$

Example 1 (Cont). $\mathsf{g}_c(\mathbf{Q}_0) = \{\{b_2\}\}$ since $\texttt{Nec}_{pr}(a, \mathbb{G}_0) = \{\{(b_2, a)\}\}$.

Example 2 (Cont). Recall that $\mathbf{Q}_i = \langle \text{pr}, \mathbb{G}_i, a \rangle$ with $i = 1, ..., 6$.

- $\mathsf{g}_c(\mathbf{Q}_1) = \{\{b_1\}, \{b_2\}\}$
- $\mathsf{g}_c(\mathbf{Q}_2) = \{\{b_2\}\}$
- $\mathsf{g}_c(\mathbf{Q}_3) = \emptyset$
- $\mathsf{g}_c(\mathbf{Q}_4) = \{\{b_1, b_2\}, \{b_1, b_3\}, \{b_2, b_3\}\}$
- $\mathsf{g}_c(\mathbf{Q}_5) = \{\{b_1\}\}$
- $\mathsf{g}_c(\mathbf{Q}_6) = \{\{b\}\}$

Remarks: Explanations provided by the functions g_f and g_c **inherit** all properties of sufficient sets and necessary sets respectively. Indeed, g_f **guarantees at least one** explanation for every query while g_c does not. The empty set can be returned by g_f but not by g_c. Furthermore, despite the fact that the two functions answer different questions, their explanations are closely related. We show (consequence of Theorem 13) that their explanations are **dual**. Indeed, every explanation provided by g_f is a **minimal hitting set** of all explanations provided by g_c for the same argument, and vice versa.

Theorem 15. *Let* $\mathbf{Q} = \langle \delta, \mathbb{G}, a \rangle$ *be a query where* $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$*.*

- $E \in \mathsf{g}_f(\mathbf{Q})$ *iff E is a subset-minimal of attackers of a such that* $\forall E' \in \mathsf{g}_c(\mathbf{Q})$*,* $E \cap E' \neq \emptyset$*.*
- $E \in \mathsf{g}_c(\mathbf{Q})$ *iff E is a subset-minimal of attackers of a such that* $\forall E' \in \mathsf{g}_f(\mathbf{Q})$*,* $E \cap E' \neq \emptyset$*.*

The query $\mathbf{Q}_1 = \langle \text{pr}, \mathbb{G}_1, a \rangle$ has two counterfactual explanations: $\{b_1\}$ and $\{b_2\}$. Note that the removal of the attack coming from b_i results in credulous acceptance (i.e., $\texttt{Acc}_{pr}(a, \mathbb{G}_1 \ominus \{(b_i, a)\}) = \texttt{c}$). One may rather be interested in knowing how to alter the status of a to sceptical acceptance. Our last function answers such questions.

[**Q3:**] ***Why not*** *the status* $\texttt{s}$ *instead of x for argument a from graph $\mathbb{G}$ under semantics δ?*

The so-called *contrastive* explanations in the XAI literature answer such questions [31]. They show how to get **another desirable outcome** (status $\texttt{s}$ in our case) than the current one. The example of $\mathbf{Q}_1$ shows that removing *one* necessary set is not sufficient. The query $\mathbf{Q}_5$ shows that removing *all* necessary sets does not answer the question either since, due to the **inhibited attack** from b_2, a would be undecided. Theorem 5 shows that to be sceptically accepted, all influential attacks, including the inhibited ones, should be removed. Hence, the next function returns the sources of all influential attacks.

Definition 16 (g_t)**.** *A* contrastive explainer *is a function* g_t *such that for every query* $\mathbf{Q} = \langle \delta, \mathbb{G}, a \rangle$*, with* $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle$*,*

$$\mathsf{g}_t(\mathbf{Q}) = \{\{b \in \texttt{Args} \mid (b, a) \in \texttt{Inf}_\delta(a, \mathbb{G})\}\}.$$

This function returns a single explanation since an argument has only one set of influential attacks.

Property 1. *For any query* $\mathbf{Q}$*,* $|\mathsf{g}_t(\mathbf{Q})| = 1$*.*

Example 1 (Cont). $\mathsf{g}_t(\mathbf{Q}_0) = \mathsf{g}_c(\mathbf{Q}_0) = \{\{b_2\}\}$.

Example 2 (Cont). Recall that $\mathbf{Q}_i = \langle \text{pr}, \mathbb{G}_i, a \rangle$ with $i = 1, \ldots, 6$.

- $\mathsf{g}_t(\mathbf{Q}_1) = \{\{b_1, b_2\}\}$
- $\mathsf{g}_t(\mathbf{Q}_2) = \{\{b_2\}\}$
- $\mathsf{g}_t(\mathbf{Q}_3) = \{\emptyset\}$
- $\mathsf{g}_t(\mathbf{Q}_4) = \{\{b_1, b_2, b_3\}\}$
- $\mathsf{g}_t(\mathbf{Q}_5) = \{\{b_1, b_2\}\}$
- $\mathsf{g}_t(\mathbf{Q}_6) = \{\{a, b\}\}$

From Theorem 14, it follows that any factual explanation of g_f and any counterfactual of g_c is included in the contrastive explanation provided by g_t.

Property 2. *Let* $x \in \{f, c\}$ *and* $\mathbf{Q}$ *a query. If* $E \in \mathsf{g}_x(\mathbf{Q})$*, then* $E \subseteq E'$ *where* $E' \in \mathsf{g}_t(\mathbf{Q})$*.*

In the XAI literature, a key criterion of goodness of an explanation is its **size**. The shorter an explanation, the better it is [31] as it would easily be grasped. The size of explanations provided by the three novel functions is bounded by the number of attackers of the discussed argument. They are thus shorter and simpler than explanations based on sub-graphs, defenders, or admissibility.

Proposition 16. *Let* $\mathbf{Q} = \langle \delta, \mathbb{G}, a \rangle$ *be a query and* $x \in \{f, c, t\}$*. For any* $E \in \mathsf{g}_x(\mathbf{Q})$*,* $0 \leq |E| \leq \texttt{Att}(a, \mathbb{G})$*.*

5 Conclusion

The paper investigates explainability of acceptability status of arguments under extension semantics. Contrary to most of the existing works, which adopt a model-based approach for providing tailored explanations to each acceptability status based on the basic notions of admissibility and defense, this paper adopts a post-hoc approach to provide a general method for selecting insights to explain the acceptability status of arguments. It considers a semantics as a black-box and looks for correlations between argumentation graphs and acceptability status assigned to their arguments by the semantics. The local direct attackers of an argument provide *sufficient*, *counterfactual*, and *contrastive* explanations of its acceptability status. The approach is general as it explains in a unified way any acceptability status and under any extension semantics which satisfies the two properties of monotonicity and maximality, including thus the four classical semantics from [22]. It is worth mentioning that the notions of sufficiency and necessity have been used in [11] for explaining extension semantics, but they are based on defenders.

This work lends itself to a number of developments including the study of explanation of semantics in richer argumentation settings (eg., bipolar [14], constrained [16], incomplete[30]).

6 Appendix

6.1 Background on Labellings

In [12] another approach for defining extension semantics has been proposed. It consists of assigning labels ($\texttt{In}, \texttt{Out}, \texttt{Und}$) to arguments.

Definition 17 (Complete Labelling)**.** *Let $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$. A complete labelling is a total function $\mathcal{L} : \mathcal{A} \to \{\texttt{In}, \texttt{Out}, \texttt{Und}\}$ such that for any $a \in \mathcal{A}$,*

1. $\mathcal{L}(a) = \texttt{In}$ *iff* $\forall (b, a) \in \texttt{Att}(a, \mathbb{G}), \mathcal{L}(b) = \texttt{Out}$.
2. $\mathcal{L}(a) = \texttt{Out}$ *iff* $\exists (b, a) \in \texttt{Att}(a, \mathbb{G})$ *such that* $\mathcal{L}(b) = \texttt{In}$.

Notations: Let $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$. We denote by $\texttt{Lab}_{co}(\mathbb{G})$ the set of all complete labellings of $\mathbb{G}$. For $X \subseteq \mathcal{A}$, we denote $\mathcal{L} = \texttt{Ext2Lab}(X) \in \texttt{Lab}(\mathbb{G})$ such that:

- $\forall x \in X, \mathcal{L}(x) = \texttt{In}$,
- $\forall y \in \mathcal{A} \setminus X$ such that $\exists x \in X$ and $x \mathcal{R} y$, $\mathcal{L}(y) = \texttt{Out}$,
- $\forall y \in \mathcal{A} \setminus X$ such that $\nexists x \in X$ and $x \mathcal{R} y$, $\mathcal{L}(y) = \texttt{Und}$.

For a labelling $\mathcal{L}$ of $\mathbb{G}$, we denote by $\texttt{Lab2Ext}(\mathcal{L})$ a function which returns the arguments labelled $\texttt{In}$, i.e. $\texttt{Lab2Ext}(\mathcal{L}) = \{x \in \mathcal{A} \mid \mathcal{L}(x) = \texttt{In}\}$.

Proof. of Property 1. Follows from Proposition 4 and Definition 16. □

Proof. of Property 2. Follows from definitions 14 and 15 and Theorem 14. □

Property 3 ([12])**.** *Let $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$.*

- *For any $\mathcal{L} \in \texttt{Lab}_{\texttt{co}}(\mathbb{G})$, $\texttt{Lab2Ext}(\mathcal{L}) \in \texttt{Ext}_{co}(\mathbb{G})$.*
- *For any $\mathcal{E} \in \texttt{Ext}_{co}(\mathbb{G})$, $\texttt{Ext2Lab}(\mathcal{E}) \in \texttt{Lab}_{co}(\mathbb{G})$.*

6.2 Proofs of Lemmas

We provide below some results that are useful for our proofs.

Lemma 17. *Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, and $a \in \mathcal{A}$. For any $X \in \texttt{Suff}_\delta(a, \mathbb{G})$ such that $X \neq \emptyset$, it holds that*

$$\texttt{Nec}_\delta(a, \mathbb{G} \downarrow X) = \{\{b_1\}, \ldots, \{b_k\}\},$$

where $X = \{b_1, \ldots, b_k\}$.

Proof. Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$ and $X \in \texttt{Suff}_\delta(a, \mathbb{G})$ such that $X \neq \emptyset$. Let $X = \{b_1, \ldots, b_k\}$. By definition of $\texttt{Suff}$, it holds that:

i) $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X)$ and
ii) $\forall Y \subset X, \texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \downarrow Y)$.

Let $b \in X$. From ii), $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, (\mathbb{G} \downarrow X) \ominus \{b\})$. From i), $\texttt{Acc}_\delta(a, \mathbb{G} \downarrow X) \neq \texttt{Acc}_\delta(a, (\mathbb{G} \downarrow X) \ominus \{b\})$. From Proposition 7, $\emptyset$ cannot be a necessary set, then $\{b\} \in \texttt{Nec}_\delta(a, \mathbb{G} \downarrow X)$. □

Lemma 18. *Let δ be a monotonic extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, and $a \in \mathcal{A}$ such that $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$. For any $X \subseteq \texttt{Att}(a, \mathbb{G})$ such that $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X)$, it holds that $\forall Y \in \texttt{Nec}_\delta(a, \mathbb{G} \downarrow X)$, $\exists Z \in \texttt{Nec}_\delta(a, \mathbb{G})$ with $Y \cap Z \neq \emptyset$.*

Proof. Let δ be a monotonic extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, and $a \in \mathcal{A}$ such that $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$. Let $X \subseteq \texttt{Att}(a, \mathbb{G})$ such that $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X)$ (1). From the first property of Theorem 2, $\forall T \subseteq \texttt{Att}(a, \mathbb{G})$,

$\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow (X \cup T))$ (2).

Assume now that $Y \in \texttt{Nec}(a, \mathbb{G} \downarrow X)$. So, $\texttt{Acc}_\delta(a, \mathbb{G} \downarrow X) \neq \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X \setminus Y)$. From (1), $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X \setminus Y)$. Let $Z = \texttt{Att}(a, \mathbb{G}) \setminus X$. Then, $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus (Z \cup Y))$. From Proposition 6, $\exists T \subseteq Z \cup Y$ such that $T \in \texttt{Nec}(a, \mathbb{G})$. Then, $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus T)$. Assume that $T \subseteq Z$. Note that $\mathbb{G} \ominus T = \mathbb{G} \downarrow (X \cup T')$ where $T' = Z \setminus T$. From (2), $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow (X \cup T'))$. contradiction. □

Lemma 19. *Let $\delta \in \{gr, co, pr, st\}$, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$ and $a, b \in \mathcal{A}$. If $\texttt{Att}(a, \mathbb{G}) = \texttt{Att}(b, \mathbb{G})$, then $\forall \mathcal{L} \in \mathbf{L}(\mathbb{G})$, $\mathcal{L}(a) = \mathcal{L}(b)$.*

Proof. Let $\delta \in \{\text{gr, co, pr, st}\}$, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$. Let also $a, b \in \mathcal{A}$ such that $\texttt{Att}(a, \mathbb{G}) = \texttt{Att}(b, \mathbb{G})$ and $\mathcal{L} \in \mathbf{L}(\mathbb{G})$. If $\texttt{Att}(a, \mathbb{G}) = \texttt{Att}(b, \mathbb{G}) = \emptyset$, then $\mathcal{L}(a) = \mathcal{L}(b) = \texttt{In}$ (see Def. 17). Assume now that $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$. There are three cases:

- $\exists c \in \texttt{Att}(a, \mathbb{G})$ such that $\mathcal{L}(c) = \texttt{In}$. From Def. 17, $\mathcal{L}(a) = \mathcal{L}(b) = \texttt{Out}$ (as $c \in \texttt{Att}(b, \mathbb{G})$).
- $\forall c \in \texttt{Att}(a, \mathbb{G})$, $\mathcal{L}(c) = \texttt{Out}$. From Def. 17, $\mathcal{L}(a) = \mathcal{L}(b) = \texttt{In}$ (since $c \in \texttt{Att}(b, \mathbb{G})$).
- Else, $\mathcal{L}(a) = \mathcal{L}(b) = \texttt{Und}$.

□

6.3 Proofs of Propositions

Proof. of Proposition 3. Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$ and $X \subseteq \texttt{Dum}_\delta(a, \mathbb{G})$.

▶ If $\texttt{Att}(a, \mathbb{G}) = \emptyset$, then $X = \texttt{Dum}_\delta(a, \mathbb{G}) = \emptyset$ and so $\mathbb{G} = \mathbb{G} \ominus X$. Consequently, $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X)$.

▶ Assume that $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$. If $\texttt{Dum}_\delta(a, \mathbb{G}) = \emptyset$, then $X = \emptyset$, $\mathbb{G} = \mathbb{G} \ominus X$ and $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X)$.

Assume now that $\texttt{Dum}_\delta(a, \mathbb{G}) \neq \emptyset$. Let $X = \{b_1, \ldots, b_k\}$, $X_0 = \emptyset$ and $X_i = \{b_1, \ldots, b_i\}$, hence

$$\forall\, 1 \leq i \leq k,\ X_i = X_{i-1} \cup \{b_i\}.$$

For any $b_i \in X$, since b_i is dummy, then $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X_{i-1}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X_{i-1} \cup \{b_i\})$. So,

$$\forall i \leq k,\ \ \texttt{Acc}_\delta(a, \mathbb{G} \ominus X_{i-1}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X_i).$$

Hence, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X_0) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X_k)$ while $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X_0) = \texttt{Acc}_\delta(a, \mathbb{G})$ and $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X_k) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X)$. □

Proof. of Proposition 4. Straightforward from the definitions of dummy and influential. □

Proof. of Proposition 6. Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$ such that $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$, and $X \subseteq \texttt{Att}(a, \mathbb{G})$. Assume that $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus X)$. Either $X \in \texttt{Nec}_\delta(a, \mathbb{G})$ or $X \notin \texttt{Nec}_\delta(a, \mathbb{G})$ meaning that X is not minimal for set-inclusion. Thus, $\exists X' \subset X$ such that $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus X')$. □

Proof. of Proposition 7. Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, and $a \in \mathcal{A}$. Note that $\mathbb{G} \ominus \emptyset = \mathbb{G}$, hence $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus \emptyset)$. This means that $\emptyset$ is not necessary for a. □

Proof. of Proposition 8. Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$, and $X \subseteq \texttt{Att}(a, \mathbb{G})$. If $\texttt{Att}(a, \mathbb{G}) = \emptyset$, then $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow \emptyset)$, hence $\emptyset \in \texttt{Suff}_\delta(a, \mathbb{G})$. Assume now that $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$ and $X \subseteq \texttt{Att}(a, \mathbb{G})$. Assume also that $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X)$. Either $X \in \texttt{Suff}_\delta(a, \mathbb{G})$ or $X \notin \texttt{Suff}_\delta(a, \mathbb{G})$ meaning that X is not minimal for set-inclusion. Thus, $\exists X' \subset X$ such that $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X')$. We repeat the same reasoning with X'. □

Proof. of Proposition 9. Let δ be an extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, and $a \in \mathcal{A}$. Assume that $\texttt{Suff}_\delta(a, \mathbb{G}) = \emptyset$. Thus, $\forall X \subseteq \texttt{Att}(a, \mathbb{G})$, X is not sufficient for a. Thus, $\texttt{Att}(a, \mathbb{G})$ is not sufficient for a, i.e., $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus \emptyset)$. This is impossible since $\mathbb{G} = \mathbb{G} \ominus \emptyset$. □

Proof. of Proposition 16. Straightforward from the definitions. □

6.4 Proofs of Theorems

Proof. of Theorem 1. Maximality has been shown in [3] and Monotonicity in [5]. □

Proof. of Theorem 2. Let δ be a monotonic semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$ and $a \in \mathcal{A}$.

Assume that $\texttt{Att}(a, \mathbb{G}) = \emptyset$. Hence, for any $X \subseteq \texttt{Att}(a, \mathbb{G})$, $X = \emptyset$. Thus, $\mathbb{G} \downarrow X = \mathbb{G}$ and the first property holds. The condition of the second property is not satisfied.

Assume now that $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$ and let $X \subseteq \texttt{Att}(a, \mathbb{G})$.

Let $\texttt{Att}_\mathbb{G}(a) = X \cup Y \cup Z$ and $X' = X \cup Y$. Note that $\mathbb{G} \downarrow X' = \mathbb{G} \ominus Z$, $\mathbb{G} \downarrow X = (\mathbb{G} \downarrow X') \ominus Y$. Since δ is monotonic, then $\texttt{Acc}_\delta(a, \mathbb{G} \downarrow X) \succeq \texttt{Acc}_\delta(a, \mathbb{G} \ominus Z) \succeq \texttt{Acc}_\delta(a, \mathbb{G})$, hence $\texttt{Acc}_\delta(a, \mathbb{G} \downarrow X) \succeq \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X') \succeq \texttt{Acc}_\delta(a, \mathbb{G})$. Assume that $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X)$. Then, $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X')$.

Assume that $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus X)$ (A1). Let $\texttt{Att}(a, \mathbb{G}) = X \cup Y \cup Z$ with $X \cap Y = \emptyset$. From monotony of δ, $\texttt{Acc}_\delta(a, \mathbb{G}) \succeq \texttt{Acc}_\delta(a, \mathbb{G} \ominus X) \succeq \texttt{Acc}_\delta(a, \mathbb{G} \ominus (X \cup Y))$. If $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus (X \cup Y))$, then $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X)$, which contradicts the assumption (A1) □

Proof. of Theorem 5. Let δ be an extension semantics which satisfies satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$ and $a \in \mathcal{A}$ such that $\texttt{Inf}_\delta(a, \mathbb{G}) \neq \emptyset$.

▶ The inequality $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{s}$ follows from Theorem 10. Indeed, since $\texttt{Inf}_\delta(a, \mathbb{G}) \neq \emptyset$ and from Proposition 4, $\texttt{Att}(a, \mathbb{G}) = \texttt{Dum}_\delta(a, \mathbb{G}) \cup \texttt{Inf}_\delta(a, \mathbb{G})$, then $\texttt{Dum}_\delta(a, \mathbb{G}) \neq \texttt{Att}(a, \mathbb{G})$ and so $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{s}$.

▶ Let us show that $\texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Inf}_\delta(a, \mathbb{G})) = \texttt{s}$.

By definition, an argument is either dummy of influential. Assume $\texttt{Dum}_\delta(a, \mathbb{G}) = \emptyset$. Hence, $\texttt{Inf}_\delta(a, \mathbb{G}) = \texttt{Att}(a, \mathbb{G})$. So $\texttt{Att}(a, \mathbb{G} \ominus \texttt{Inf}_\delta(a, \mathbb{G})) = \emptyset$ and $\texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Inf}_\delta(a, \mathbb{G})) = \texttt{s}$ (from Maximality).

Let now $\texttt{Dum}_\delta(a, \mathbb{G}) = \{b_1, \ldots, b_k\}$, $X_0 = \texttt{Inf}_\delta(a, \mathbb{G})$ and $X_i = \texttt{Inf}_\delta(a, \mathbb{G}) \cup \{b_1, \ldots, b_i\}$. Since elements of $\texttt{Dum}_\delta(a, \mathbb{G})$ are dummy, then:

$$\begin{aligned}
&\texttt{Acc}_\delta(a, \mathbb{G} \ominus X_0) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X_1)\\
&\texttt{Acc}_\delta(a, \mathbb{G} \ominus X_1) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X_2)\\
&\qquad\qquad\vdots\\
&\texttt{Acc}_\delta(a, \mathbb{G} \ominus X_{k-1}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X_k)
\end{aligned}$$

Hence, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X_0) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X_k)$. From Proposition 4, $X_k = \texttt{Att}(a, \mathbb{G})$. So, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Inf}_\delta(a, \mathbb{G})) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Att}(a, \mathbb{G}))$. Since $\texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Att}(a, \mathbb{G})) = \texttt{s}$, then $\texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Inf}_\delta(a, \mathbb{G})) = \texttt{s}$. □

Proof. of Theorem 10. Let δ be a monotonic extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$ and $a \in \mathcal{A}$. Note that if $\texttt{Att}(a, \mathbb{G}) = \emptyset$, then $\texttt{Dum}_\delta(a, \mathbb{G}) = \texttt{Att}(a, \mathbb{G}) = \emptyset$. Moreover, the unique subset of $\texttt{Att}(a, \mathbb{G})$ is $X = \emptyset$ and $\mathbb{G} = \mathbb{G} \ominus X$. Thus, $\texttt{Suff}_\delta(a, \mathbb{G}) = \{\emptyset\}$. Finally, from Maximality, $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{s}$.

Let now $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$.

▶ Assume that $\texttt{Ext}_\delta(\mathbb{G}) = \emptyset$. From Def.3, $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{s}$. From monotonicity of δ, $\forall X \subseteq \texttt{Att}(a, \mathbb{G})$, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) \succeq \texttt{Acc}_\delta(a, \mathbb{G})$. Hence, $\forall X \subseteq \texttt{Att}(a, \mathbb{G})$, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) = \texttt{s}$. This means that $\texttt{Dum}_\delta(a, \mathbb{G}) = \texttt{Att}(a, \mathbb{G})$. Note also that $\texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Att}(a, \mathbb{G})) = \texttt{s}$, hence $\emptyset \in \texttt{Suff}_\delta(a, \mathbb{G})$. Assume that $X \in \texttt{Suff}_\delta(a, \mathbb{G})$. X being minimal by definition, it follows that $X = \emptyset$. So, $\texttt{Suff}_\delta(a, \mathbb{G}) = \{\emptyset\}$.

▶ Assume that $\texttt{Ext}_\delta(\mathbb{G}) \neq \emptyset$.

Assume that $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{s}$. From monotonicity of δ, $\forall X \subseteq \texttt{Att}(a, \mathbb{G})$, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) \succeq \texttt{Acc}_\delta(a, \mathbb{G})$. Hence, $\forall X \subseteq \texttt{Att}(a, \mathbb{G})$, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) = \texttt{s}$. Thus, $\texttt{Dum}_\delta(a, \mathbb{G}) = \texttt{Att}(a, \mathbb{G})$. Furthermore, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Att}(a, \mathbb{G})) = \texttt{s}$. This means that $\emptyset \in \texttt{Suff}_\delta(a, \mathbb{G})$. Assume that $X \in \texttt{Suff}_\delta(a, \mathbb{G})$. X being minimal by definition, it follows that $X = \emptyset$. Then, $\texttt{Suff}_\delta(a, \mathbb{G}) = \{\emptyset\}$.

Assume now that $\texttt{Suff}_\delta(a, \mathbb{G}) = \{\emptyset\}$. Then, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Att}(a, \mathbb{G})) = \texttt{Acc}_\delta(a, \mathbb{G})$. In the graph $\mathbb{G} \ominus \texttt{Att}(a, \mathbb{G})$, the argument a is not attacked. Maximality of δ ensures that $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{s}$. From monotonicity of δ, $\forall X \subseteq \texttt{Att}(a, \mathbb{G})$, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) \succeq \texttt{Acc}_\delta(a, \mathbb{G})$. Hence, $\forall X \subseteq \texttt{Att}(a, \mathbb{G})$, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) = \texttt{s}$. Thus, $\texttt{Dum}_\delta(a, \mathbb{G}) = \texttt{Att}(a, \mathbb{G})$.

Assume that $\texttt{Dum}_\delta(a, \mathbb{G}) = \texttt{Att}(a, \mathbb{G})$. Hence, $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Att}(a, \mathbb{G}))$. Maximality of δ ensures that $\texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Att}(a, \mathbb{G})) = \texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{s}$. Following the reasoning above, $\texttt{Suff}_\delta(a, \mathbb{G}) = \{\emptyset\}$. □

Proof. of Theorem 11. Let δ be an extension semantics that satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$.

Assume that $\texttt{Suff}_\delta(a, \mathbb{G}) = \{\emptyset\}$. It follows that $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus \texttt{Att}(a, \mathbb{G}))$. Let $\mathbb{G}' = \mathbb{G} \ominus \texttt{Att}(a, \mathbb{G})$. Since a has no attackers in $\mathbb{G}'$, then from Maximality $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G}') = \texttt{s}$. Assume now that some $X \in \texttt{Nec}_\delta(a, \mathbb{G})$, so

$$\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus X) \qquad \text{(C1)}.$$

From monotonicity of δ, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) \succeq \texttt{Acc}_\delta(a, \mathbb{G})$. So, $\texttt{Acc}_\delta(a, \mathbb{G} \ominus X) = \texttt{s}$, which contradicts (C1).

Assume that $\texttt{Nec}_\delta(a, \mathbb{G}) = \emptyset$. Then, $\forall X \subseteq \texttt{Att}(a, \mathbb{G})$, $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \ominus X)$. Thus, $\emptyset \in \texttt{Suff}_\delta(a, \mathbb{G})$. By definition, sufficient sets are subset-minimal, then if $X \in \texttt{Suff}_\delta(a, \mathbb{G})$, then $X = \emptyset$. □

Proof. of Theorem 13. Let δ be an extension semantics that satisfies Maximality and Monotonicity, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$ and $a \in \mathcal{A}$ such that $\texttt{Att}(a, \mathbb{G}) \neq \emptyset$.

▶ Let us now show the first property.

Let $X \in \texttt{Suff}_\delta(a, \mathbb{G})$, hence $\texttt{Acc}_\delta(a, \mathbb{G}) = \texttt{Acc}_\delta(a, \mathbb{G} \downarrow X)$ **(A1)**. If $X = \emptyset$, then the property is satisfied in a vacuous way since from Theorem 11, $\texttt{Nec}_\delta(a, \mathbb{G}) = \emptyset$. Assume that $X \neq \emptyset$, then from Theorem 11, $\texttt{Nec}_\delta(a, \mathbb{G}) \neq \emptyset$. Let $Y \in \texttt{Nec}_\delta(a, \mathbb{G})$, so $\texttt{Acc}_\delta(a, \mathbb{G}) \neq \texttt{Acc}_\delta(a, \mathbb{G} \ominus Y)$ **(A2)**. Assume that $X \cap Y = \emptyset$. Let $\texttt{Att}(a, \mathbb{G}) = X \cup Y \cup Z$. Note that $\mathbb{G} \ominus Y = \mathbb{G} \downarrow (X \cup Z)$. From

(A1) and Theorem 2, it follows that $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow (X \cup Z)) = \texttt{Acc}_\delta(a,\mathbb{G} \ominus Y)$, which contradicts **(A2)**.

Let $X \subseteq \texttt{Att}(a,\mathbb{G})$ be such that X is a minimal (for set inclusion) set which satisfies the condition: $\forall Y \in \texttt{Nec}_\delta(a,\mathbb{G})$, $X \cap Y \neq \emptyset$ (A1). Let us show that $X \in \texttt{Suff}_\delta(a,\mathbb{G})$. We first show that $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow X)$. Let $\texttt{Nec}_\delta(a,\mathbb{G}) = \{Y_1,\ldots,Y_k\}$ such that $\forall i = 1,\ldots,k$, $X \cap Y_i \neq \emptyset$. Let $\texttt{Att}(a,\mathbb{G}) = X \cup Z$ such that $X \cap Z = \emptyset$. Thus, $\forall i = 1,\ldots,k$, $Y_i \not\subseteq Z$ **(A2)**. Assume that $\texttt{Acc}_\delta(a,\mathbb{G}) \neq \texttt{Acc}_\delta(a,\mathbb{G} \downarrow X)$, i.e., $\texttt{Acc}_\delta(a,\mathbb{G}) \neq \texttt{Acc}_\delta(a,\mathbb{G} \ominus Z)$. From Proposition 6, $\exists T \subseteq Z$ such that $T \in \texttt{Nec}_\delta(a,\mathbb{G})$, which contradicts **(A2)**. So, $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow X)$.

Let us now show the minimality of X. Assume $\exists X' \subset X$ such that $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow X')$ **(A1)**. Since X is the minimal set verifying the condition **(A1)**, then X' violates the condition meaning that $\exists Y \in \texttt{Nec}_\delta(a,\mathbb{G})$ such that $X' \cap Y = \emptyset$. Note that $\texttt{Acc}_\delta(a,\mathbb{G}) \neq \texttt{Acc}_\delta(a,\mathbb{G} \ominus Y)$ **(A2)**. Let $\texttt{Att}(a,\mathbb{G}) = X' \cup Y \cup Z$. Note that $\mathbb{G} \ominus Y = \mathbb{G} \downarrow (X' \cup Z)$. From Theorem 2 and **(A1)**, it follows that $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow (X' \cup Z))$, which contradicts **(A2)**.

► Let us now show the second property.

Let $X \in \texttt{Nec}_\delta(a,\mathbb{G})$, then $\texttt{Acc}_\delta(a,\mathbb{G}) \neq \texttt{Acc}_\delta(a,\mathbb{G} \ominus X)$ (A1). From Theorem 11, $\emptyset \notin \texttt{Suff}_\delta(a,\mathbb{G})$. Let $Y \in \texttt{Suff}_\delta(a,\mathbb{G})$, so $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow Y)$ (A2). Assume that $X \cap Y = \emptyset$. Let $\texttt{Att}(a,\mathbb{G}) = X \cup Y \cup Z$. Note that $\mathbb{G} \ominus X = \mathbb{G} \downarrow (Y \cup Z)$. From (A2) and first property in Theorem 2, it follows that $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow (X \cup Z)) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow X)$, which contradicts (A1).

Let now X be a minimal (for set inclusion) subset of $\texttt{Att}(a,\mathbb{G})$ such that $\forall Y \in \texttt{Suff}_\delta(a,\mathbb{G})$, $X \cap Y \neq \emptyset$ (A1). Let us show that $X \in \texttt{Nec}_\delta(a,\mathbb{G})$. By definition of X, it follows that $X \neq \emptyset$ and $\emptyset \notin \texttt{Suff}_\delta(a,\mathbb{G})$. From Theorem 11, $\texttt{Nec}_\delta(a,\mathbb{G}) \neq \emptyset$. Let $\texttt{Att}(a,\mathbb{G}) = X \cup Z$ and $\texttt{Suff}_\delta(a,\mathbb{G}) = \{Y_1,\ldots,Y_k\}$. Note that due to (A1), $\forall i = 1,\ldots,k$, $Y_i \not\subseteq Z$ (A2). Assume that $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \ominus X)$. Note that $\mathbb{G} \ominus X = \mathbb{G} \downarrow Z$. From Proposition 8, $\exists T \subseteq Z$ such that $T \in \texttt{Suff}_\delta(a,\mathbb{G})$, which contradicts (A2). Assume now that X violates minimality, i.e., $\exists X' \subset X$ such that $\texttt{Acc}_\delta(a,\mathbb{G}) \neq \texttt{Acc}_\delta(a,\mathbb{G} \ominus X')$. Since X is the minimal set satisfying (A1), then $\exists Y \in \texttt{Suff}_\delta(a,\mathbb{G})$ such that $X' \cap Y = \emptyset$. Furthermore, $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow Y)$. Let $\texttt{Att}(a,\mathbb{G}) = X' \cup Y \cup Z$. Note that $\mathbb{G} \ominus X' = \mathbb{G} \downarrow (Y \cup Z)$. From Theorem 2, $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow (Y \cup Z))$. So, $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \ominus X')$, which is a contradiction. Hence, $X \in \texttt{Nec}_\delta(a,\mathbb{G})$. □

Proof. of Theorem 14. Let δ be a monotonic extension semantics, $\mathbb{G} = \langle \mathcal{A}, \mathcal{R} \rangle \in \texttt{AG}$, $a \in \mathcal{A}$ and $b \in \texttt{Att}(a,\mathbb{G})$.

► Let us show the property 2.

Let us show the inclusion $\bigcup_{X \in \texttt{Suff}_\delta(a,\mathbb{G})} X \subseteq \bigcup_{Y \in \texttt{Nec}_\delta(a,\mathbb{G})} Y$.

If $\texttt{Suff}(a,\mathbb{G}) = \{\emptyset\}$, then from Theorem 11, $\texttt{Nec}(a,\mathbb{G}) = \emptyset$. So, $\bigcup_{X \in \texttt{Suff}_\delta(a,\mathbb{G})} X = \bigcup_{Y \in \texttt{Nec}_\delta(a,\mathbb{G})} Y = \emptyset$.

Assume now $\texttt{Suff}(a,\mathbb{G}) \neq \{\emptyset\}$ and let $X \in \texttt{Suff}(a,\mathbb{G})$ and $b \in X$. It follows that: $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow X)$ and $\forall Y \subset X$, $\texttt{Acc}_\delta(a,\mathbb{G}) \neq \texttt{Acc}_\delta(a,\mathbb{G} \downarrow Y)$. From Lemma 17, $\texttt{Nec}_\delta(a,\mathbb{G} \downarrow X) = \{\{b_i\} \mid b_i \in X\}$, so $\{b\} \in \texttt{Nec}_\delta(a,\mathbb{G} \downarrow X)$. From Lemma 18, $\exists Z \in \texttt{Nec}_\delta(a,\mathbb{G})$ such that $Z \cap \{b\} \neq \emptyset$, so $b \in \bigcup_{Y \in \texttt{Nec}_\delta(a,\mathbb{G})} Y$.

Let us show the inclusion $\bigcup_{X \in \texttt{Nec}_\delta(a,\mathbb{G})} X \subseteq \bigcup_{Y \in \texttt{Suff}_\delta(a,\mathbb{G})} Y$.

Let $b \in X$ where $X \in \texttt{Nec}_\delta(a,\mathbb{G})$. Then, $\texttt{Acc}_\delta(a,\mathbb{G}) \neq \texttt{Acc}_\delta(a,\mathbb{G} \ominus X)$ (1) and $\forall Y \subset X$, $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \ominus Y)$ (2). Let $Z = \texttt{Att}(a,\mathbb{G}) \setminus X$. Then, from (2) $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow (Z \cup \{b\}))$. From Proposition 8, $\exists Z' \subseteq Z \cup \{b\}$ such that $Z' \in \texttt{Suff}_\delta(a,\mathbb{G})$, i.e., $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow Z')$. Assume $Z' \subseteq Z$. From Theorem 2, $\texttt{Acc}_\delta(a,\mathbb{G}) = \texttt{Acc}_\delta(a,\mathbb{G} \downarrow Z) = \texttt{Acc}_\delta(a,\mathbb{G} \ominus X)$, which contradicts (1). Hence, $b \in Z'$ and so $b \in \bigcup_{Y \in \texttt{Suff}_\delta(a,\mathbb{G})} Y$.

► Assume that b is dummy and $\exists X \in \texttt{Nec}(a,\mathbb{G})$ such that $b \in X$. Then,

$$\texttt{Acc}_\delta(a,\mathbb{G}) \neq \texttt{Acc}_\delta(a,\mathbb{G} \ominus X) \quad (1)$$

Let $X = Z \cup \{b\}$. Since b is dummy, then

$$\texttt{Acc}_\delta(a,\mathbb{G} \ominus Z) = \texttt{Acc}_\delta(a,\mathbb{G} \ominus X) \quad (2)$$

From (1), $\texttt{Acc}_\delta(a,\mathbb{G} \ominus Z) \neq \texttt{Acc}_\delta(a,\mathbb{G})$, which contradicts the minimality of X. Hence, $b \in \texttt{Inf}(a,\mathbb{G})$. □

Proof. of Theorem 15. Straightforward from Theorem 13. □

Acknowledgements

This work was supported by the AI Interdisciplinary Institute ANITI, funded by the French program "AI-Cluster".

References

[1] L. Amgoud. Explaining black-box classifiers: Properties and functions. *Int. J. Approx. Reason.*, 155:40–65, 2023. doi: 10.1016/J.IJAR.2023.01.004. URL https://doi.org/10.1016/j.ijar.2023.01.004.

[2] L. Amgoud. Post-hoc explanation of extension semantics, 2024.

[3] L. Amgoud and J. Ben-Naim. Axiomatic foundations of acceptability semantics. In *Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning KR*, pages 2–11, 2016.

[4] L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173:413–436, 2009.

[5] L. Amgoud, J. Ben-Naim, and S. Vesic. Measuring the intensity of attacks in argumentation graphs with shapley value. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI*, pages 63–69, 2017.

[6] K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G. R. Simari, M. Thimm, and S. Villata. Towards artificial argumentation. *AI Magazine*, 38(3):25–36, 2017.

[7] P. Baroni and M. Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15): 675–700, 2007.

[8] P. Baroni, D. Gabbay, M. Giacomin, and L. van der Torre, editors. *Handbook of Formal Argumentation, Volume 1*. College Publications, 2018.

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

[10] E. Bonzon and N. Maudet. On the outcomes of multiparty persuasion. In *10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011), Taipei, Taiwan, May 2-6, 2011, Volume 1-3*, pages 47–54, 2011.

[11] A. Borg and F. Bex. Minimality, necessity and sufficiency for argumentation and explanation. *International Journal of Approximate Reasoning*, 168:109143, 2024.

[12] M. Caminada. On the issue of reinstatement in argumentation. In *Proceedings of the 10th European Conference on Logics in Artificial Intelligence JELIA'06*, pages 111–123, 2006.

[13] C. Cayrol and M. Lagasquie-Schiex. Graduality in argumentation. *Journal of Artificial Intelligence Research*, 23:245–297, 2005.

[14] C. Cayrol and M. Lagasquie-Schiex. Bipolar abstract argumentation systems. In G. R. Simari and I. Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 65–84. Springer, 2009.

[15] C. Cayrol, S. Doutre, and J. Mengin. Dialectical proof theories for the credulous preferred semantics of argumentation frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU*, volume 2143 of *Lecture Notes in Computer Science*, pages 668–679. Springer, 2001.

[16] S. Coste-Marquis, C. Devred, and P. Marquis. Constrained argumentation frameworks. In *Proceedings, Tenth International Conference on Principles of Knowledge Representation and Reasoning*, pages 112–122, 2006.

[17] K. Cyras, D. Letsios, R. Misener, and F. Toni. Argumentation for explainable scheduling. In *The Thirty-Third Conference on Artificial Intelligence, AAAI*, pages 2752–2759, 2019.

[18] J. Delobelle and S. Villata. Interpretability of gradual semantics in abstract argumentation. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 15th European Conference, ECSQARU*, volume 11726 of *Lecture Notes in Computer Science*, pages 27–38. Springer, 2019.

[19] Y. Dimopoulos, J. Mailly, and P. Moraitis. Argumentation-based negotiation with incomplete opponent profiles. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS*, pages 1252–1260, 2019.

[20] S. Doutre, T. Duchatelle, and M. Lagasquie-Schiex. Classes of explanations for the verification problem in abstract argumentation. In *17èmes Journées d'Intelligence Artificielle Fondamentale, JIAF 2023*, pages 124–134, 2023.

[21] S. Doutre, T. Duchatelle, and M. Lagasquie-Schiex. Visual explanations for defence in abstract argumentation. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, AAMAS*, pages 2346–2348. ACM, 2023.

[22] P. M. Dung. On the Acceptability of Arguments and its Fundamental Role in Non-Monotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence*, 77:321–357, 1995.

[23] X. Fan and F. Toni. On explanations for non-acceptable arguments. In *Third International Workshop on Theory and Applications of Formal Argumentation, TAFA*, volume 9524 of *Lecture Notes in Computer Science*, pages 112–127. Springer, 2015.

[24] X. Fan and F. Toni. On computing explanations in argumentation. In *Proceedings of the Twenty-Ninth Conference on Artificial Intelligence, AAAI*, pages 1496–1502. AAAI Press, 2015.

[25] A. J. García, C. I. Chesñevar, N. D. Rotstein, and G. R. Simari. Formalizing dialectical explanation support for argument-based reasoning in knowledge-based systems. *Expert Syst. Appl.*, 40(8):3233–3247, 2013.

[26] D. Grossi and S. Modgil. On the graded acceptability of arguments. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI'15*, pages 868–874, 2015.

[27] N. Hadidi, Y. Dimopoulos, and P. Moraitis. Argumentative alternating offers. In *9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS*, pages 441–448, 2010.

[28] T. Kampik, K. Cyras, and J. R. Alarcón. Change in quantitative bipolar argumentation: Sufficient, necessary, and counterfactual explanations. *International Journal of Approximate Reasoning*, 164:109066, 2024.

[29] B. Liao and L. van der Torre. Explanation semantics for abstract argumentation. In *Computational Models of Argument - Proceedings of COMMA*, volume 326, pages 271–282. IOS Press, 2020.

[30] J. Mailly. Extension-based semantics for incomplete argumentation frameworks: properties, complexity and algorithms. *Journal of Logic and Computation*, 33(2):406–435, 2023.

[31] T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019.

[32] S. Modgil and M. Caminada. *Proof Theories and Algorithms for Abstract Argumentation Frameworks*, pages 105–129. Springer US, 2009.

[33] N. Potyka, X. Yin, and F. Toni. Explaining random forests using bipolar argumentation and markov networks. In *Thirty-Seventh Conference on Artificial Intelligence, AAAI*, pages 9453–9460. AAAI Press, 2023.

[34] C. Sakama. Abduction in argumentation frameworks and its use in debate games. In Y. I. Nakano, K. Satoh, and D. Bekki, editors, *New Frontiers in Artificial Intelligence - JSAI-isAI 2013 Workshops, LENLS, JURISIN, MiMI, AAA, and DDS*, volume 8417 of *Lecture Notes in Computer Science*, pages 285–303. Springer, 2013.

[35] Z. G. Saribatur, J. P. Wallner, and S. Woltran. Explaining non-acceptability in abstract argumentation. In *24th European Conference on Artificial Intelligence, ECAI*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 881–888, 2020.

[36] G. Simari, M. Giacomin, D. Gabbay, and M. Thimm, editors. *Handbook of Formal Argumentation, Volume 2*. College Publications, 2021.

[37] G. R. Simari and I. Rahwan, editors. *Argumentation in Artificial Intelligence*. Springer, 2009. doi: 10.1007/978-0-387-98197-0.

[38] M. Ulbricht and J. P. Wallner. Strong explanations in abstract argumentation. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*, pages 6496–6504. AAAI Press, 2021.

[39] H. Wachsmuth, M. Potthast, K. Al Khatib, Y. Ajjour, J. Puschmann, J. Qu, J. Dorsch, V. Morari, J. Bevendorff, and B. Stein. Building an argument search engine for the web. In *Proceedings of the 4th Workshop on Argument Mining*, pages 49–59, 2017.

[40] Z. Zeng, C. Miao, C. Leung, Z. Shen, and J. J. Chin. Computing argumentative explanations in bipolar argumentation frameworks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):10079–10080, 2019.