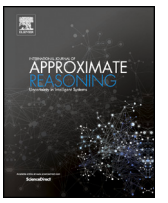
International Journal of Approximate Reasoning 168 (2024) 109143

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar



Minimality, necessity and sufficiency for argumentation and explanation

AnneMarie Borg [a,*], Floris Bex [a,b,*]

[a] Department of Information and Computing Sciences, Utrecht University, The Netherlands
[b] Tilburg Institute for Law, Technology, and Society, Tilburg University, The Netherlands

ARTICLE INFO

Keywords:
Formal argumentation
Structured argumentation frameworks
Explainable artificial intelligence
AI and law

ABSTRACT

We discuss explanations for formal (abstract and structured) argumentation – the question whether and why a certain argument or claim can be accepted (or not) under various extension-based semantics. We introduce a flexible framework, which can act as the basis for many different types of explanations. For example, we can have simple or comprehensive explanations in terms of arguments for or against a claim, arguments that (indirectly) defend a claim, the evidence (knowledge base) that supports or is incompatible with a claim, and so on. We show how selection based on necessity and sufficiency can be captured in our basic framework and discuss a real-life application.

1. Introduction

Explainable Artificial Intelligence (XAI) is an important and fast growing research area, which now also incorporates findings from the humanities and social sciences on how humans request, generate, interpret and evaluate explanations – see, for example, [42] and [4], and [46] for an extensive survey. Argumentation, more specifically formal argumentation [8,30], plays an important role in XAI, in various different ways. For example, there is a conceptual link between argumentation and explanation as different forms of reasoning [1,11]. Furthermore, formal argumentation has been used to explain the output of other, less interpretable AI models (see [24] for a recent overview). Finally, with formal argumentation increasingly being used in real-world applications [3], there has also been an increase in work on explaining the output of the argument-based systems themselves [16,17,33,34,36,40,61,63]. In this paper, we focus on the latter connection between argumentation and XAI: how can we explain the conclusions of formal argumentation systems, taking into account some of the ideas from the literature on how humans select and evaluate explanations [46]?

A central concept in formal argumentation is that of *abstract argumentation frameworks* [30], sets of arguments and the attack relations between them. For such an abstract argumentation framework we can determine extensions, sets of arguments that can collectively be considered as accepted under different semantics [30]. What we are interested in here is explaining the (non-)acceptance of a particular argument in an argumentation framework, in other words, an answer to the question 'why is argument A (not) accepted given the set of arguments and attacks in the argumentation framework?' Using just the basic extensions of Dung [30] we could answer this question with 'because A belongs to (does not belong to) a set of accepted arguments, an extension'. However, as

* Corresponding authors.
E-mail addresses: a.borg@uu.nl (A. Borg), f.j.bex@uu.nl (F. Bex).

https://doi.org/10.1016/j.ijar.2024.109143
Received 5 September 2023; Received in revised form 22 December 2023; Accepted 7 February 2024

Available online 12 February 2024

already noted by Fan and Toni [33], this answer does not explain which arguments are *relevant* for the (non-)acceptance of argument A, as the extensions that contain A may (partly) consist of arguments that have no influence on the acceptability of A at all. Rather, we would like to have an explanation in terms of relevant arguments, that is, arguments that influence the acceptability of A by attacking A or defending A (i.e., attacking attackers of A). For example, as an explanation for the acceptance of A we could have 'A is accepted because argument C defends A against its only attacking argument B', and as an explanation for the non-acceptance of A we could have 'A is not accepted because it is attacked by B and C, which are both accepted'.

In addition to the abstract argumentation frameworks of Dung [30], formal argumentation is also about *structured* or *logical argumentation frameworks* [9], where arguments are constructed from a knowledge base and a set of rules, and the attack relation is based on the individual elements in the arguments (e.g., an argument with conclusion p attacks an argument with conclusion $\neg p$ and vice versa). With respect to explanations, the explanandum (that which must be explained) then becomes the (non-)acceptance of a formula, that is, the conclusion of some argument instead of the (non-)acceptance of an abstract argument. That explanations for arguments and explanations for formulas can differ is clear if we consider that, for example, some formula ϕ may be the conclusion of multiple different arguments, some of which may be accepted and some of which may not. Furthermore, in structured argumentation frameworks we can also give explanations *in terms of* formulas - for example, 'formula ϕ is accepted because it is the conclusion of an accepted argument with premises ψ and χ'.

The first main aim of this paper is then to provide a generic, flexible formal framework for determining and computing argument-based explanations for arguments and their conclusions. Such an explanation provides the relevant reasons for (not) accepting an argument or conclusion under different semantics and for different types of reasoners (e.g., credulous or skeptical reasoners). Though work on such explanations exists in the literature (see e.g., [16,33,34,36,40,61,63]), our framework is *generic* in that the underlying argumentation framework does not have to be adjusted and the definitions are semantics-independent. For example, the explanations based on the new semantics of Fan and Toni [33] are a special case of our framework. Building on our earlier work in [16], we provide new results on the properties of explanations in Section 3.3, add an algorithm for computing explanations in Section 3.4, and show in Section 4.2 how explanations can be adjusted to account for naïve semantics, stage semantics, and special cases under admissible and stable semantics. The framework is also *flexible*, as the contents of explanations can be varied according to different definitions of relevance and taking into account different elements of arguments. Again building on [16], we expand the discussion of different types of explanations in this paper by investigating formal properties in Section 4.3 and providing a realistic example based on an argumentation system in use at the Netherlands Police [54] in Section 7.

Our second aim is to take into account how humans select relevant explanations based on minimality, necessity and sufficiency [46]. Simpler explanations that contain fewer elements (in our case: relevant arguments) are usually preferred [62], and one way to ensure this is to select the most (subset-)minimal explanation. For example, in the case of the above explanation for the non-acceptance of A, we could also say 'A is not accepted because it is attacked by B, which is accepted', providing a more minimal explanation than the one mentioning both attackers B and C. Two other important criteria for explanation selection are *necessity* and *sufficiency* [42]. In the context of (abstract) argumentation, an argument B is sufficient for the (non-)acceptance of A if no other arguments than B are needed for A to (not) be accepted, and B is necessary for the (non-)acceptance of A if B is always needed for the (non-)acceptance of A. Again returning to our example explanation for not accepting A, we can say that both B and C individually are sufficient for the non-acceptance of A (if A is attacked by only B or C it is still not accepted), and neither B nor C is necessary (if either B or C is not there, the other argument still attacks A). For structured argumentation, the explanations can again be given in terms of formulas. It may, for example, be the case that in our example both B and C have the formula ψ as a premise, making ψ necessary for the non-acceptance of A.

In this paper we will study how the above minimal, necessary and sufficient explanations can be constructed, given an argumentation framework under different semantics. We will show in Section 6 how minimal explanations can be selected from the explanations recalled in Section 3 and show how selection based on necessity and sufficiency can be integrated. This discussion is more elaborate and general than in [17]: we will provide more formal details (i.e., we formally define minimal explanations, provide extensive motivation for our definitions of necessity and sufficiency, especially in the ASPIC$^+$ setting where we also add existence explanations and we make a distinction between credulous and skeptical (non-)acceptance) and in our extended example of the police argumentation system in Section 7 show how necessity and sufficiency are especially useful to reduce the size of an explanation to the core reason(s) for a conclusion.

The paper is structured as follows. In Section 2 we provide the preliminaries on abstract argumentation and some of the used notions after which we present the basic framework of explanations in Section 3, which also contains a polynomial time algorithm to compute the explanations from an argumentation framework and its extensions. Then, in Section 4, we discuss several variations of the basic explanations and properties of the resulting explanations. In Section 5 explanations for structured argumentation settings are introduced. Necessary and sufficient explanations for both the abstract and structured setting are studied in Section 6 and the real-life example is studied in Section 7. Section 8 contains an extensive discussion on related literature and we conclude in Section 9.

2. Preliminaries: abstract argumentation

In this section we recall the most important notions from abstract argumentation [30] and introduce some additional notions necessary for the basic framework for explanations. The preliminaries on structured argumentation will be provided in Section 5.1.

An *abstract argumentation framework* (AF) [30] is a pair $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$, where Args is a set of *arguments* and $\mathsf{Att} \subseteq \mathsf{Args} \times \mathsf{Args}$ is an *attack relation* on these arguments. An argumentation framework can be viewed as a directed graph, in which the nodes represent arguments and the arrows represent attacks between arguments. See Fig. 1 for an example.

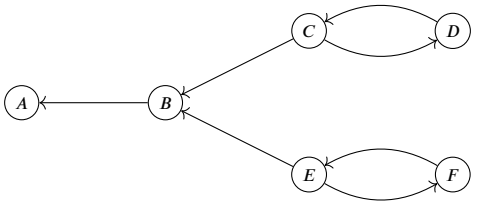


Fig. 1. Graphical representation of the argumentation framework $\mathcal{AF}_1$.

Example 1. Fig. 1 represents the argumentation framework $\mathcal{AF}_1 = \langle \mathsf{Args}_1, \mathsf{Att}_1 \rangle$ where $\mathsf{Args}_1 = \{A, B, C, D, E, F\}$ and $\mathsf{Att}_1 = \{(B, A), (C, B), (C, D), (D, C), (E, B), (E, F), (F, E)\}$.

Given an argumentation framework $\mathcal{AF}$, Dung-style semantics [30] can be applied to it, to determine what combinations of arguments (called *extensions*) can collectively be accepted. We consider here many of the most applied semantics, see [5] for a detailed discussion and additional semantics.

Definition 1. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, $\mathsf{S} \subseteq \mathsf{Args}$ a set of arguments and let $A \in \mathsf{Args}$. Then:

- S *attacks* A if there is an $A' \in \mathsf{S}$ such that $(A', A) \in \mathsf{Att}$, let S^+ denote the set of arguments attacked by S;
- S *defends* A if S attacks every attacker of A;
- S is *conflict-free* if there are no $A_1, A_2 \in \mathsf{S}$ such that $(A_1, A_2) \in \mathsf{Att}$;
- S is a *naïve extension* (Nav) if it is a maximal (with respect to $\subseteq$) conflict-free set;
- S is a *stage extension* (Stg) of $\mathcal{AF}$ if $\mathsf{S} \cup \mathsf{S}^+$ is maximal (with respect to $\subseteq$) among the conflict-free sets; and
- S is an *admissible extension* (Adm) if it is conflict-free and it defends all of its elements.
- An admissible extension that contains all the arguments that it defends is a *complete extension* (Cmp).
- The *grounded extension* (Grd) is the minimal (with respect to $\subseteq$) complete extension;
- A *preferred extension* (Prf) is a maximal (with respect to $\subseteq$) complete extension;
- An *ideal extension* (Idl) is the maximal (with respect to $\subseteq$) admissible set that is included in each preferred extension;
- A *stable extension* (Stb) is a complete extension that attacks every argument not in it;
- A *semi-stable extension* (Sstb) is a complete extension for which $\mathsf{S} \cup \mathsf{S}^+$ is maximal (with respect to $\subseteq$); and
- An *eager extension* (Egr) is the maximal (with respect to $\subseteq$) admissible set that is included in every semi-stable extension.

$\mathsf{Sem}(\mathcal{AF})$ denotes the set of all the extensions of $\mathcal{AF}$ under the semantics $\mathsf{Sem} \in \{\mathsf{Nav}, \mathsf{Stg}, \mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$. For single extension semantics (i.e., $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Idl}, \mathsf{Egr}\}$) we will sometimes identify $\mathsf{Sem}(\mathcal{AF})$ with its single element.

Example 2. For the argumentation framework $\mathcal{AF}_1$ from Example 1 and shown in Fig. 1 we have that the grounded extension is empty (i.e., $\mathsf{Grd}(\mathcal{AF}_1) = \{\emptyset\}$) and there are four preferred, stable and semi-stable extensions: $\{A, C, E\}$, $\{A, C, F\}$, $\{A, D, E\}$ and $\{B, D, F\}$ (i.e., $\mathsf{Sem}(\mathcal{AF}_1) = \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4\} = \{\{A, C, E\}, \{A, C, F\}, \{A, D, E\}, \{B, D, F\}\}$ for $\mathsf{Sem} \in \{\mathsf{Prf}, \mathsf{Stb}, \mathsf{Sstb}\}$).

With the following notation we can collect all the extensions that (do not) contain a certain argument under a given semantics.

Notation 1. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, $A \in \mathsf{Args}$ and $\mathsf{S} \subseteq \mathsf{Args}$. Then, for some $\mathsf{Sem} \in \{\mathsf{Nav}, \mathsf{Stg}, \mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$:

- $\mathsf{SemWith}(A) = \{\mathcal{E} \in \mathsf{Sem}(\mathcal{AF}) \mid A \in \mathcal{E}\}$ denotes the set of Sem-extensions of $\mathcal{AF}$ that contain A;
- $\mathsf{SemWithout}(A) = \{\mathcal{E} \in \mathsf{Sem}(\mathcal{AF}) \mid A \notin \mathcal{E}\}$ denotes the set of Sem-extensions of $\mathcal{AF}$ that do not contain A.

Example 3. For the argumentation framework $\mathcal{AF}_1$ from Example 1 we have that $\mathsf{PrfWith}(A) = \{\{A, C, E\}, \{A, C, F\}, \{A, D, E\}\}$ and $\mathsf{PrfWith}(B) = \{\{B, D, F\}\}$ while $\mathsf{PrfWithout}(D) = \{\{A, C, E\}, \{A, C, F\}\}$ and $\mathsf{PrfWithout}(E) = \{\{A, C, F\}, \{B, D, F\}\}$.

The next definition introduces the *acceptance strategies*. In addition to choosing a semantics with which an argumentation framework is evaluated, a user can choose how to (not) accept an argument: skeptically (i.e., only accepting arguments that are always accepted) or credulously (i.e., accepting arguments as soon as this is possible).[1]

Definition 2. Where $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ is an argumentation framework and Sem a semantics with $\mathsf{Sem}(\mathcal{AF}) \neq \emptyset$, it is said that $A \in \mathsf{Args}$ is:

[1] In e.g., [16] arguments that are not accepted could be *skeptically non-accepted* or *credulously non-accepted*. Since this causes some confusion on the intuition of what *skeptical* [resp. *credulous*] *non-acceptance* means, we will use here the notions *not skeptically* [resp. *credulously*] *accepted*.

- *skeptically accepted* if $\mathsf{SemWith}(A) = \mathsf{Sem}(\mathcal{AF})$, that is: A is part of all Sem-extensions;
- *credulously accepted* if $\mathsf{SemWith}(A) \neq \emptyset$, that is: A is part of at least one Sem-extension.

We will denote by $\cap$ the skeptical acceptance strategy and by $\cup$ the credulous acceptance strategy. When the strategy is arbitrary, results in the same (i.e., there is no difference between $\cap$ and $\cup$, such as when $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Idl}, \mathsf{Egr}\}$) or is clear from the context, we will refer to *accepted* arguments, leaving out the exact acceptance strategy.

In what follows we will often refer to arguments that are not skeptically or not credulously accepted: arguments that are not part of all Sem-extensions (not skeptically accepted) or arguments that are not part of any Sem-extension (not credulously accepted). Like for accepted arguments, when clear from the context we will refer to non-accepted arguments.

Example 4. Recall that $\mathsf{Sem}(\mathcal{AF}_1) = \{\{A, C, E\}, \{A, C, F\}, \{A, D, E\}, \{B, D, F\}\}$ for $\mathsf{Sem} \in \{\mathsf{Prf}, \mathsf{Stb}, \mathsf{Sstb}\}$. It follows that none of the arguments are skeptically accepted, but all arguments are credulously accepted for admissible, complete, preferred, stable and semi-stable semantics.

3. Basic explanations

Given an argumentation framework and some argument that is (not) (skeptically/credulously) accepted under some semantics, we are now looking for meaningful explanations of *why* an argument is (not) (skeptically/credulously) accepted under the relevant semantics. The first possibility would be to provide the extensions of which the argument is part. For example, if we ask 'Given $\mathcal{AF}_1$, why is argument A credulously accepted under Prf?' One answer could be 'Because it is part of three of the four preferred extensions (see Example 2)'. However, this explanation does not tell us the *reasons* for why A is credulously accepted, rather it clarifies how Definitions 1 and 2 lead to credulous acceptance of A in $\mathcal{AF}_1$. More specifically, it does not tell us exactly which arguments influenced or caused A's acceptance.

One of the core ideas of explanations is that they point to a relevant cause[2] C of the explanandum E: if in some hypothetical counterfactual case C is not the case, then the explanandum E will also not be the case [39]. In the case of formal argumentation, we would like to know which arguments in the argumentation framework influenced (or caused) the (non-)acceptance of the argument that makes up the explanandum. This interpretation of explanations in argumentation has become common in the literature [16,17,33,34,36,40,61,63], as it is closely related to explanations for other AI models (e.g., which features caused the model to classify this picture as a spider [46]), and hence lends itself well to a broader discussion on types of explanations and selecting them.

In formal argumentation, arguments are connected to each other via the attack relation. Based on this relation, there are two ways for an argument B to influence the acceptability of another argument A in an argumentation framework: B can *attack* or *defend* (attack an attacker of A). Note that attack or defense does not have to be direct: for example, in $\mathcal{AF}_1$ the acceptance of argument D can also influence that of A, since D attacks C, which defends A against its (direct) attacker B. So in the context of explanations we are interested in both *direct* and *indirect* attack and defense between arguments. In Definition 1, only direct attack was defined (i.e., Att), and defense was only defined between a set of arguments and an argument. Below, we therefore define direct and indirect attack and defense between arguments.

Definition 3. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, $A, B \in \mathsf{Args}$ and $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for some Sem.

We say that A *defends* B if A directly or indirectly defends B, where: A *directly defends* B if there is some $C \in \mathsf{Args}$ such that $(C, B) \in \mathsf{Att}$ and $(A, C) \in \mathsf{Att}$, and A *indirectly defends* B if A (directly or indirectly) defends $C \in \mathsf{Args}$ and C (directly or indirectly) defends B. Similarly, we say that A *attacks* B if A directly or indirectly attacks B, where: A *directly attacks* B if $(A, B) \in \mathsf{Att}$ and A *indirectly attacks* B if A (directly or indirectly) attacks some $C \in \mathsf{Args}$ and C (directly or indirectly) defends B. It is said that A *defends* B *in* $\mathcal{E}$ if A (directly or indirectly) defends B and $A \in \mathcal{E}$.

Now, direct and indirect attack and defense are used to define when one argument is *relevant* for another argument.

Definition 4. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and let $A, B \in \mathsf{Args}$. It is said that A is *relevant* for B if A (in)directly attacks or defends B and it does not attack itself. A set $\mathsf{S} \subseteq \mathsf{Args}$ is relevant for B if all of its arguments are relevant for B.

Example 5. In the argumentation framework $\mathcal{AF}_1$ from Example 1, the argument F attacks E directly and A indirectly, and defends both B and F directly. The arguments D and F are relevant for A and B but not relevant for each other.

The intuition behind relevance is that for an argument B to be relevant to A, changing the acceptance of B can (potentially) influence the (non-)acceptability of argument A.[3] In our example $\mathcal{AF}_1$, all of the arguments in $\mathsf{Args}_1 \setminus \{A\}$ are relevant for A, since

[2] Note that here we are not talking about any kind of physical causation or causation in the real world. Rather, as is fairly common in XAI, we consider there to be a causal influence relation between the input and output of an AI model.

[3] Taking as an analogy traditional accounts of explanation [46], one could say that the status of B can (potentially) *cause* a change in the status of A.

changing the acceptance of any of them could influence the acceptability of A. On the other hand, changing the acceptability of A or D will never influence the acceptability of F, so they are not relevant for F. For an even clearer example, consider $\mathcal{AF}_1$ with one argument G added to Args_1 and no further new attacks. Since G is not connected to any of the other arguments via attacks, changing the acceptance of G will never influence the acceptance of any of the other arguments, and thus G is not relevant for any of the other arguments.

Note that the notions of attack and defense, and hence the notion of relevance, are defined independently of semantics, that is, of the extensions of an argumentation framework. So, for example, in $\mathcal{AF}_1$ arguments F and A are both part of the preferred extension $\{A, C, F\}$ even though F indirectly attacks A. Similarly, an argument having defenders does not mean that the argument is in some extension, or that the argument and its defenders will always be in the same extension(s). For example, consider the argumentation framework $\mathcal{AF} = \langle\{A, B, C\}, \{(A, B), (B, C), (C, A)\}\rangle$. Even though A is defended by arguments A, B and C, A is not in any complete extension of $\mathcal{AF}$. This is why we say that for relevance, an argument could *potentially* influence the acceptability of another argument – whether it *actually* influences the acceptability depends on the semantics used and the possible extensions under these semantics. In the following sections, this will become more clear when we relate relevance to particular extensions in order to construct explanations for (non-)acceptance.

In the rest of this section the basic explanations for accepted (Section 3.1) and non-accepted arguments (Section 3.2) are introduced. Properties of these explanations are then discussed in Section 3.3 and an algorithm to compute the explanations is presented in Section 3.4. Once these basic explanations are formally defined, we will discuss variations in Section 4, their application in ASPIC$^+$ in Section 5 and ways to select smaller sets of arguments in terms of the notions of minimality, necessity and sufficiency in Section 6.

3.1. Basic explanations for acceptance

Now that we have introduced a basic notion of relevance, we can start answering the question for acceptance explanations: given an argumentation framework $\mathcal{AF}$ and some argument A that is (skeptically/credulously) accepted under semantics Sem, what are the relevant arguments that explain *why* A is (skeptically/credulously) accepted under Sem?

Admissibility-based semantics (i.e., Adm, Cmp, Grd, Prf, Idl, Stb, Sstb and Egr) are all based on a notion of defense (cf. Definition 1). It therefore makes sense for explanations to consider as the relevant arguments those arguments that successfully defend A against its attackers in the extension(s) under these semantics. For this, we need to collect all defenders of an argument (cf. Definition 3), and all defenders of an argument that are in a particular extension, as per the following definition.

Definition 5. Let $\mathcal{AF} = \langle\mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework, $A \in \mathsf{Args}$ and $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ an extension for some semantics Sem.

- $\mathsf{Defending}(A) = \{B \in \mathsf{Args} \mid B \text{ defends } A\}$;
- $\mathsf{Defending}(A, \mathcal{E}) = \mathsf{Defending}(A) \cap \mathcal{E}$. This denotes the set of arguments that defend A in $\mathcal{E}$;

We now define two types of acceptance explanations for each semantics Sem, based on the skeptical ($\cap$) and credulous ($\cup$) acceptance strategies.[4] The $\cap$-explanations provide all the reasons why an argument can be accepted by a skeptical reasoner: for each extension it contains the set of arguments that defend the argument.[5] The $\cup$-explanations provide one reason why an argument can be accepted by a credulous reasoner: for a given extension it contains the set of arguments that defend the argument.[6]

Definition 6 *(Acceptance argument explanation).* Let $\mathcal{AF} = \langle\mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework, let $A, B \in \mathsf{Args}$ and $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ a semantics. Suppose that B is skeptically accepted and that B is credulously accepted. Moreover, let $\mathcal{E}_A \in \mathsf{SemWith}(A)$ be some Sem-extension containing A. Then:

$$\mathsf{SemAcc}^{\cap}(B) = \{\mathsf{Defending}(B, \mathcal{E}) \mid \mathcal{E} \in \mathsf{SemWith}(A)\};$$

$$\mathsf{SemAcc}^{\cup}(A, \mathcal{E}_A) = \mathsf{Defending}(A, \mathcal{E}_A).$$

Thus, $\mathsf{SemAcc}^{\cap}(B)$ is a set that contains for each Sem-extension the set of arguments that defend the skeptically accepted argument B in that extension (why is B skeptically accepted under Sem or, in other words, which arguments defend B in each extension?), and $\mathsf{SemAcc}^{\cup}(A, \mathcal{E}_A)$ provides the set of arguments that defend A in the Sem-extension $\mathcal{E}_A$ (why is A credulously accepted under Sem or, in other words, which arguments defend A in extension $\mathcal{E}_A$?). Since the grounded extension is unique we will often identify $\mathsf{GrdAcc}^{\cap}$ with its single element and we will not distinguish between $\cap$ and $\cup$. Note that a skeptical acceptance explanation for an

[4] We assume the explanandum is true, that is, an explanation for the acceptance of an argument A is only requested when A is accepted with respect to the considered semantics and acceptance strategy.

[5] In [16], the $\cap$-explanation was the union of all these sets. In this paper, we define it as a set of sets. This makes the defenders of A for all the different Sem-extensions explicit.

[6] In [16], the extension in the $\cup$-explanation was not explicit, but random. In this paper, we make the extension explicit. This gives a user the possibility to further specify the scenario of the explanation while in an implementation it would still be possible to provide a random extension.

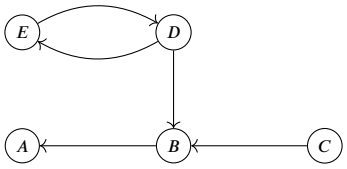


Fig. 2. Graphical representation of the argumentation framework $\mathcal{AF}_2$.

argument A can only be requested when A is skeptically accepted, while a credulous acceptance explanation for an argument B can be requested when B is credulously accepted.

Example 6. Consider the argumentation framework $\mathcal{AF}_1$ from Example 1. From Example 4 it is known that none of the arguments from Args_1 are skeptically accepted, thus $\mathsf{SemAcc}^{\cap}$ cannot be requested for any of the considered semantics. However, we do have that:

- $\mathsf{Prf}\,\mathsf{Acc}^{\cup}(A, \mathcal{E}_1) = \{C, E\}$, $\mathsf{Prf}\,\mathsf{Acc}^{\cup}(A, \mathcal{E}_2) = \{C\}$ and $\mathsf{Prf}\,\mathsf{Acc}^{\cup}(A, \mathcal{E}_3) = \{E\}$;
- $\mathsf{Prf}\,\mathsf{Acc}^{\cup}(B, \mathcal{E}_4) = \{D, F\}$.

So A is credulously accepted under Prf because it is defended by C and E in extension $\mathcal{E}_1 = \{A, C, E\}$, by C in extension $\mathcal{E}_2 = \{A, C, F\}$ and by E in extension $\mathcal{E}_3 = \{A, D, E\}$. B is credulously accepted under Prf because it is defended by D and F in extension $\mathcal{E}_4 = \{B, D, F\}$.

In order to illustrate the $\mathsf{SemAcc}^{\cap}$-explanation, we introduce the following example.

Example 7. Let $\mathcal{AF}_2$ be the argumentation framework as shown in Fig. 2. In this framework we have that $\mathsf{Sem}(\mathcal{AF}_2) = \{\mathcal{E}_1, \mathcal{E}_2\} = \{\{A, C, D\}, \{A, C, E\}\}$ for $\mathsf{Sem} \in \{\mathsf{Prf}, \mathsf{Stb}, \mathsf{Sstb}\}$ and $\mathsf{Sem}(\mathcal{AF}_2) = \{\mathcal{E}_3\} = \{\{A, C\}\}$ for $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Idl}, \mathsf{Egr}\}$. Here, we have the following explanations:

- $\mathsf{GrdAcc}(A) = \{C\}$;
- $\mathsf{Prf}\,\mathsf{Acc}^{\cap}(A) = \{\{C\}, \{C, D\}\}$;
- $\mathsf{Prf}\,\mathsf{Acc}^{\cup}(A, \mathcal{E}_1) = \{C, D\}$ and $\mathsf{Prf}\,\mathsf{Acc}^{\cup}(A, \mathcal{E}_2) = \{C\}$.

So A is skeptically accepted under Grd because it is defended by C in the single grounded extension $\mathcal{E}_3$, and A is skeptically and credulously accepted under Prf because it is defended by either C (in extension $\mathcal{E}_2 = \{A, C, E\}$) or by C and D (in extension $\mathcal{E}_1 = \{A, D, E\}$).

3.2. Basic explanations for non-acceptance

In addition to acceptance explanations, we also want to be able to explain non-acceptance: given an argumentation framework $\mathcal{AF}$ and some argument A that is not (skeptically/credulously) accepted under semantics Sem, what are the relevant arguments that explain *why* A is not (skeptically/credulously) accepted under Sem?

In any completeness-based semantics (e.g., Grd, Cmp, Prf or Sstb), an argument is not accepted if it is attacked and it is not defended by an accepted argument. Hence, intuitively, the explanation for the non-acceptance of an argument is the set of attacking arguments for which no defense exists. For this, we need to collect all attackers of an argument (cf. Definition 3) for which there is no defender in a particular extension.

Definition 7. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, $A \in \mathsf{Args}$ and $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ an extension for some semantics Sem.

- $\mathsf{NoDef\,Against}(A, \mathcal{E}) = \{B \in \mathsf{Args} \mid B \text{ attacks } A \text{ and } \mathcal{E} \text{ does not attack } B\}$. This denotes the set of all attackers of A for which no defense exists from $\mathcal{E}$.

We can then define two types of non-acceptance explanations. The $\cap$-explanation provides one reason for why an argument cannot be accepted by a skeptical reasoner: for a given extension it contains the set of arguments against which there is no defense. The $\cup$-explanations provide all the reasons for why an argument cannot be accepted by a credulous reasoner: for each extension it contains the set of arguments against which there is no defense.

Definition 8 *(Non-acceptance argument explanation).* Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, let $A, B \in \mathsf{Args}$ and $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Sstb}, \mathsf{Egr}\}$ a semantics. Suppose that B is not credulously accepted and that A is not skeptically accepted. Moreover, let $\mathcal{E}_A \in \mathsf{SemWithout}(A)$. Then:

$$\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}_A) = \mathsf{NoDefAgainst}(A, \mathcal{E}_A);$$

$$\mathsf{SemNotAcc}^{\cup}(B) = \{\mathsf{NoDefAgainst}(B, \mathcal{E}) \mid \mathcal{E} \in \mathsf{SemWithout}(B)\}.$$

So $\mathsf{SemNotAcc}^{\cap}(A)$ provides a set of arguments that attack A and for which no defense exists in the Sem-extension $\mathcal{E}_A$ of which A is not a member (why is A not skeptically accepted under Sem or, in other words, against which attackers of A is there no defense in extension $\mathcal{E}_A$?). $\mathsf{SemNotAcc}^{\cup}(B)$ contains for each Sem-extension the set of arguments that attack B and for which no defense exists (why is B not credulously accepted under Sem or, in other words, against which attackers of B is there no defense in all extensions?). That for $\cap$ only one extension has to be considered follows since A is not skeptically accepted as soon as there is at least one extension of which A is not a member, while B is not credulously accepted when B is not a member of any extension. Since the grounded extension is unique we will identify $\mathsf{GrdNotAcc}^{\cup}$ with its single element and make no distinction between $\cap$ and $\cup$.[7] Note that, like in the acceptance case, a credulous non-acceptance explanation for an argument B can only be requested when B is not credulously accepted, while a skeptical non-acceptance explanation for an argument A can be requested when A is not skeptically accepted.

Example 8. For the argumentation framework $\mathcal{AF}_1$ it is known (recall Example 4) that none of the arguments is not credulously accepted for $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Prf}, \mathsf{Sstb}\}$ (i.e., all arguments are credulously accepted for these semantics). Therefore, $\mathsf{SemNotAcc}^{\cup}$ cannot be requested. However, we do have that:

- $\mathsf{GrdNotAcc}(A) = \mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}_4) = \{B, D, F\}$ this follows since $\mathsf{GrdWithout}(A) = \{\emptyset\}$ and $\mathsf{PrfWithout}(A) = \{\{B, D, F\}\}$;
- $\mathsf{GrdNotAcc}(B) = \{C, E\}$ since $\mathsf{GrdWithout}(B) = \{\emptyset\}$;
- $\mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_1) = \{C, E\}$, $\mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_2) = \{C\}$ and $\mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_3) = \{E\}$ since $\mathsf{PrfWithout}(B) = \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3\} = \{\{A, C, E\}, \{A, C, F\}, \{A, D, E\}\}$

So A is not accepted under Grd and not skeptically accepted under Prf because it is not defended against B, D and F in the extensions $\emptyset$ and $\mathcal{E}_4 = \{B, D, F\}$, respectively. Similarly, B is not accepted under Grd because it is not defended against both C and E in the grounded extension $\emptyset$, while B is not skeptically accepted under Prf because it is not defended against C, E or C and E in the extensions $\mathcal{E}_2 = \{A, C, F\}$, $\mathcal{E}_3 = \{A, D, E\}$ and $\mathcal{E}_1 = \{A, C, E\}$, respectively.

Example 9. For the argumentation framework $\mathcal{AF}_2$ from Example 7, we have that B is not skeptically or credulously accepted, and D is not skeptically accepted:

- $\mathsf{PrfNotAcc}^{\cup}(B) = \{\{C\}, \{C, D\}\}$ and $\mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_1) = \{C, D\}$ and $\mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_2) = \{C\}$;
- $\mathsf{PrfNotAcc}^{\cap}(D, \mathcal{E}_2) = \{E\}$.

B is not (skeptically or credulously) accepted under Prf because in extension $\mathcal{E}_2 = \{A, C, E\}$ it is not defended against C and in extension $\mathcal{E}_1 = \{A, C, D\}$ it is not defended against C and D. D is not skeptically accepted under Prf because there is the extension $\mathcal{E}_2 = \{A, C, E\}$ in which it is not defended against E. There is no explanation $\mathsf{PrfNotAcc}^{\cup}(D)$ (why is D not credulously accepted?) because D is credulously accepted (Definition 8).

Remark 1. Non-acceptance explanations under $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Stb}\}$ are not discussed here. To see why, consider an argumentation framework with arguments A and B. Then $\{A\}$ is an admissible set, so B is not skeptically accepted, but not attacked. Now suppose that there is one attack: (B, B). Then $\{A\}$ is the only Sem-extension for $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Sstb}, \mathsf{Egr}\}$, however, there is no Stb-extension. This cannot be explained with the here presented explanations, since A does not attack B but is not attacked itself either. A solution will be discussed in Section 4.2.

3.3. Properties of the basic explanations

It is well-known (see, e.g., [5,30]) that the extensions are related to each other under different semantics. For example, every stable extension is a preferred extension and all stable and preferred extensions as well as the grounded extension are part of the set of complete extensions. Moreover, by their definition, choosing a particular semantics results in a particular type of extension. For example, by choosing the grounded semantics, one obtains an extension with arguments that very skeptical reasoners could accept, but by choosing the preferred semantics, one can obtain several extensions, some of which might include arguments that only very credulous reasoners would accept and that can be in conflict with arguments from other extensions.[8]

In this section we show how explanations under various semantics are related to each other and how acceptance and non-acceptance explanations are related. The proofs of the properties in this section can be found in A.1.

[7] In [16] the non-acceptance explanation was defined as the union of all the explanations as defined here. Here we provide different explanations that correspond to the different extensions, similar to the definition for $\mathsf{SemAcc}^{\cap}$, to better capture the notion of *not skeptically accepted* and *not credulously accepted* and their relation to each other. Similarly, we make the extension in $\mathsf{SemNotAcc}^{\cap}$ explicit, to allow the user to be more specific about the required explanation.

[8] For a detailed overview of how the extensions of different semantics are related we refer the reader to [5].

3.3.1. Properties concerning acceptance explanations

This first proposition shows how acceptance explanations are related to each other under various semantics. This is useful to know, since the type of explanation obtained from the basic framework might guide the choice of the semantics when implementing the explanations in an application.

Proposition 1. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ *be an argumentation framework where* $A \in \mathit{Args}$*. Then:*

1. $\mathsf{GrdAcc}(A) \subseteq \bigcap \mathsf{SemAcc}^{\cap}(A)$ *for all* $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$.
2. $\mathsf{StbAcc}^{\cap}(A) \subseteq \mathsf{SstbAcc}^{\cap}(A) \subseteq \mathsf{PrfAcc}^{\cap}(A) \subseteq \mathsf{CmpAcc}^{\cap}(A)$.
3. *For each* $\mathsf{S} \in \mathsf{CmpAcc}^{\cap}(A)$ *there is an* $\mathsf{S}' \in \mathsf{PrfAcc}^{\cap}(A)$ *such that* $\mathsf{S} \subseteq \mathsf{S}'$

Intuitively, the explanations obtained from the basic framework behave in a similar way as the extensions obtained under various semantics. The explanation under the grounded extension is contained in every explanation for $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ (item 1). The $\cap$-explanation for Stb is contained in the $\cap$-explanation for Sstb and Prf, which are in turn contained in the $\cap$-explanation for Cmp (item 2). As a result, all possible $\cup$-explanations for Stb are also possible $\cup$-explanations for Sstb and Prf, which are in turn possible $\cup$-explanations for Cmp. Finally, for each set in a $\cap$-explanation for Cmp there is a set in the $\cap$-explanation under Prf that contains the Cmp-explanation (item 3).

The next proposition shows how the Defending-sets of defending and defended arguments are related. This provides insight into the role of direct and indirect defenders, which will be useful when the content of the explanations is varied, as proposed in Section 4.

Proposition 2. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ *be an argumentation framework,* $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ *for* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ *and let* $A, B \in \mathit{Args}$*. Then:*

- *if* $A \in \mathsf{Defending}(B, \mathcal{E})$*, then* $\mathsf{Defending}(A, \mathcal{E}) \subseteq \mathsf{Defending}(B, \mathcal{E})$*;*
- *if* $A \in \mathsf{Defending}(B, \mathcal{E})$ *and* $B \in \mathsf{Defending}(A, \mathcal{E})$*, then* $\mathsf{Defending}(A, \mathcal{E}) = \mathsf{Defending}(B, \mathcal{E})$.

So if A is a defender of B, then the defenders of A will be a subset of the defenders of B (item 1). Furthermore, if A is a defender of B and vice versa, then the defenders of A and B are the same (item 2).

One could say that an explanation is only meaningful when it is not empty (i.e., when it actually contains something to explain the acceptance of an argument with). Otherwise, the receiver of the explanation might not be helped with their explanation-seeking question. The next proposition shows that only when an argument is not attacked its acceptance explanation will be empty.

Proposition 3. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ *be an argumentation framework and let* $A \in \mathit{Args}$ *be such that* A *is accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ *and* $\star \in \{\cap, \cup\}$*. Then* $\mathsf{SemAcc}^{\cap}(A) = \{\emptyset\}$ *and* $\mathsf{SemAcc}^{\cup}(A, \mathcal{E}) = \emptyset$ *for any* $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ *iff there is no* $B \in \mathit{Args}$ *such that* $(B, A) \in \mathit{Att}$*.*

In view of the above proposition, if the explanations are implemented into an application, the explanation method should be extended with an explanation for cases in which there are no attacks. For example, by pointing out that there are no conflicts and that, therefore, there is no defense necessary. What this explanation should look like depends on the application and the choice of the explanation (e.g., on which semantics it is based, or which explanation type (see also Section 4)).

3.3.2. Properties concerning non-acceptance explanations

Some of the properties shown above for acceptance explanations have non-acceptance counterparts as well. Recall from Remark 1 that non-acceptance explanations were not defined for $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Stb}\}$ in Section 3.2, these are therefore not included in the results. First the counterpart of Proposition 1, which shows that non-acceptance explanations behave predictable with respect to the chosen semantics as well. Thus, explanations under a certain semantics are related to explanations under another semantics in a similar ways as extensions under the first semantics are related to extensions under the other semantics.

Proposition 4. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ *be an argumentation framework where* $A \in \mathit{Args}$*. Then:*

1. $\mathsf{GrdNotAcc}(A) \subseteq \bigcap \mathsf{CmpNotAcc}^{\cup}(A)$.
2. $\mathsf{SstbNotAcc}^{\cup}(A) \subseteq \mathsf{PrfNotAcc}^{\cup}(A) \subseteq \mathsf{CmpNotAcc}^{\cup}(A)$.
3. *Let* $\mathcal{E} \in \mathsf{SstbWithout}(A)$*. Then* $\mathsf{SstbNotAcc}^{\cap}(A, \mathcal{E}) = \mathsf{SemNotAcc}^{\cap}(A, \mathcal{E})$*, for* $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Prf}\}$.

In words, the explanation for the grounded semantics is part of every Cmp-explanation (item 1) and the $\cup$-explanation for Sstb is contained in the $\cup$-explanation for Prf and Cmp (item 2). Based on this, given a Sstb-extension $\mathcal{E}$, the Sstb-explanation is also the Prf- and Cmp-explanation.

Remark 2. There is no non-acceptance counterpart of Proposition 2. Note that, for all arguments A and B and extension $\mathcal{E}$, $A \in \mathsf{NoDefAgainst}(B, \mathcal{E})$ entails that A (in)directly attacks B. Now consider the situation in which A is not accepted either. Then the

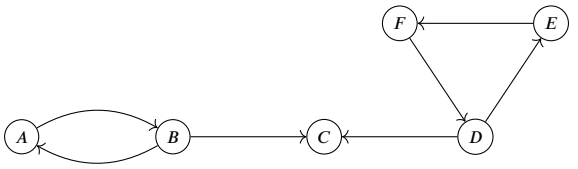


Fig. 3. Graphical representations of the argumentation framework $\mathcal{AF}_3$ from Example 10.

arguments in $\mathsf{NoDefAgainst}(A, \mathcal{E})$ (in)directly defend B (i.e., because they attack A and A attacks B). To see that such a situation can occur, take for example an argumentation framework consisting of the arguments A and B, which attack each other (i.e., (A, B) and (B, A)). Then the grounded extension is empty, $A \in \mathsf{NoDefAgainst}(B, \emptyset)$ and $B \in \mathsf{NoDefAgainst}(A, \emptyset)$, but neither argument causes its own non-acceptance.

The next proposition, the non-acceptance counterpart of Proposition 3, shows that a non-acceptance explanation is never empty. Intuitively this is the case since there has to be some reason (i.e., attacking argument) for the non-acceptance of an argument.

Proposition 5. *Let $\mathcal{AF} = \langle Args, Att \rangle$ be an argumentation framework and let $A \in Args$ be such that A is non-accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Sstb}, \mathsf{Egr}\}$ *and* $\star \in \{\cap, \cup\}$. *Then* $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) \neq \emptyset$ *for any* $\mathcal{E} \in \mathsf{SemWithout}(A)$ *and* $\mathsf{SemNotAcc}^{\cup}(A) \neq \{\emptyset\}$.

Remark 3. Proposition 5 does not hold for Adm since not every admissible extension contains all the arguments that it defends. Take for example an argumentation framework with arguments A and B and no attacks between them (Remark 1). Then $\{B\}$ is an admissible extension, thus A is not skeptically accepted, yet $\mathsf{NoDefAgainst}(A, \{B\}) = \emptyset$. In fact, depending on the extension, it could be that $\mathsf{AdmNotAcc}^{\cap}(A) = \emptyset$ and $\mathsf{AdmNotAcc}^{\cap}(B) = \emptyset$.

3.3.3. Comparing acceptance and non-acceptance

Recall from Examples 6 and 8 that acceptance and non-acceptance explanations might sometimes be very similar. In the case of these particular examples we have that $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_1) = \mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_1) = \{C, E\}$, $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_2) = \mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_2) = \{C\}$ and $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_3) = \mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_3) = \{E\}$. The next proposition shows that this is no coincidence. We show how Defending and NoDefAgainst for attacking and attacked arguments are related when either the attacked or the attacking argument(s) are part of an extension.

Proposition 6. *Let $\mathcal{AF} = \langle Args, Att \rangle$ be an argumentation framework, let $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for some* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ *and let $A, B_1, \ldots, B_n, C_1, \ldots, C_k \in Args$ such that $(B_1, A), \ldots, (B_n, A) \in Att$ and A indirectly attacks $C_1, \ldots, C_k$. Then:*

- *where $B_1, \ldots, B_m \in \mathcal{E}$, for $m \leq n$ it holds that:* $\mathsf{NoDefAgainst}(A, \mathcal{E}) \supseteq \mathsf{Defending}(B_1, \mathcal{E}) \cup \ldots \cup \mathsf{Defending}(B_m, \mathcal{E})$;
- *when $A \in \mathcal{E}$ we have:* $\mathsf{Defending}(A, \mathcal{E}) \subseteq \mathsf{NoDefAgainst}(B_1, \mathcal{E}) \cup \ldots \cup \mathsf{NoDefAgainst}(B_n, \mathcal{E})$;
- *where $A \in \mathcal{E}$ and $C_1, \ldots, C_j \notin \mathcal{E}$, for $j \leq k$ it holds that:* $\mathsf{Defending}(A, \mathcal{E}) \subseteq \mathsf{NoDefAgainst}(C_i, \mathcal{E})$ *for all* $i \in \{1, \ldots, j\}$.

The above proposition shows that the set of arguments to which an extension $\mathcal{E}$ provides no defense for A contains at least all the arguments defending the direct attackers of A in $\mathcal{E}$ (item 1). The other way around also holds: when A is part of an extension $\mathcal{E}$, it will be defended by arguments that attack its direct attackers and to which $\mathcal{E}$ provides no defense (item 2). Finally, again for an argument A part of an extension $\mathcal{E}$, the arguments that defend A contain arguments to which $\mathcal{E}$ provides no defense for arguments indirectly attacked by A (item 3).

To see that $\mathsf{NoDefAgainst}(A, \mathcal{E}) \not\subseteq \mathsf{Defending}(B_1, \mathcal{E}) \cup \ldots \cup \mathsf{Defending}(B_n, \mathcal{E})$, take a look at the following example. Intuitively this is the case since, in terms of labeling semantics [5], an argument can be `in` the extension, attacked by the extension (i.e., `out`) or attacked by an argument that is not `in` or `out` (i.e., `undecided`).

Example 10. Let $\mathcal{AF}_3 = \langle \mathsf{Args}_3, \mathsf{Att}_3 \rangle$ be the argumentation framework where $\mathsf{Args}_3 = \{A, B, C, D, E, F\}$ and Att_3 as shown in Fig. 3. There are two preferred extensions: $\mathsf{Prf}(\mathcal{AF}_3) = \{\{A\}, \{B\}\}$. Here $\mathsf{NoDefAgainst}(C, \{B\}) = \{B, D, E, F\}$, since B, D and E all attack C (in)directly but only $(B, C) \in \mathsf{Att}_3$ such that $B \in \mathcal{E}$ and $\mathsf{Defending}(B, \{B\}) = \{B\}$.

To see that acceptance and non-acceptance explanations are not necessarily exclusive, take a look at the following example.

Example 11. Let $\mathcal{AF}_4 = \langle \mathsf{Args}_4, \mathsf{Att}_4 \rangle$ be the argumentation framework where $\mathsf{Args}_4 = \{A, B, C, D, E, F\}$ and Att_4 as shown in Fig. 4. There are two preferred extensions: $\mathsf{Prf}(\mathcal{AF}_4) = \{\mathcal{E}_1, \mathcal{E}_2\} = \{\{A, B, F\}, \{B, D\}\}$. Note that both the $\cup$-acceptance and the $\cap$-non-acceptance explanations for A contain the argument B: $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_1) = \mathsf{Defending}(A, \{A, B, F\}) = \{B, F\}$, here B directly defends A against the attack from C and for non-acceptance: $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}_2) = \mathsf{NoDefAgainst}(A, \{B, D\}) = \{B, D\}$, here A is not defended against the indirect attack from B.

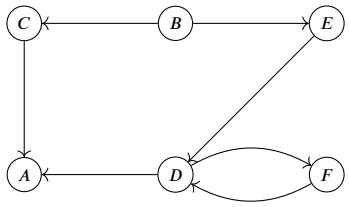


Fig. 4. Graphical representation of the argumentation framework $\mathcal{AF}_4$ from Example 11.

3.4. Computing the explanations

In order to implement the basic explanations it will be useful to have an algorithm with which the attacking and defending arguments of every argument can be determined. In this section we discuss such an algorithm. It will be shown that the algorithm is sound and complete and that, when the extensions of an argumentation framework are known (i.e., that the acceptability status of an argument is known), the algorithm runs in polynomial time.

Since an (abstract) argumentation framework can be seen as a directed graph, one can determine whether argument B is *reachable* (i.e., relevant) from argument A and if so, what the *distance* between the two arguments is. If B is reachable from A, it can be determined, based on the distance, whether A (in)directly attacks or (in)directly defends B.

Definition 9. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and let $A, B \in \mathsf{Args}$. There is an *attack-path* from A to B if $(A, B) \in \mathsf{Att}$ or there are $C_1, \ldots, C_{n-1} \in \mathsf{Args}$, such that $(A, C_1), (C_1, C_2), \ldots, (C_{n-2}, C_{n-1}), (C_{n-1}, B) \in \mathsf{Att}$ and no attack appears twice in this sequence. It is said that this attack-path has *length* n and is along the attacks $(A, C_1), (C_1, C_2), \ldots, (C_{n-1}, B)$, if $(A, B) \in \mathsf{Att}$, the path has length 1 and if $A = B$ the attack-path has length 0.[9]

Example 12. In the argumentation framework $\mathcal{AF}_1$, there are two attack-paths from C to A: $(C, B), (B, A)$ has length 2 and $(C, D), (D, C), (C, B), (B, A)$ has length 4.

Intuitively, if the length of an attack-path between arguments A and B is odd [resp. even], A (in)directly attacks [resp. defends] B. Indeed, in the example above, C directly and indirectly defends A. If we now have an algorithm to compute attack paths and their length, we can determine Defending and NoDefAgainst, and thus construct our acceptance and non-acceptance explanations.

Algorithm 1 presents a method to calculate the length of all existing attack-paths in an argumentation framework (Dist) and for each argument A the set of arguments from which a path to A exists ($\mathsf{Reach}(A)$). The algorithm is based on Procedure ReReach (*Recursive Reach*), a depth-first search algorithm.

Algorithm 1: Computing Reach and Dist.

Input : $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$
Output: For each $A, B \in \mathsf{Args}$: $\mathsf{Reach}(A)$ and $\mathsf{Dist}(A, B)$
for $A \in \mathit{Args}$ **do**
  $\mathsf{Reach}(A) = \{A\}$ and $\mathsf{Dist}(A, A) = \{0\}$;
  for $B \in \mathit{Args} \setminus \{A\}$ **do**
    $\mathsf{Dist}(A, B) = \emptyset$;
for $A \in \mathit{Args}$ **do**
  ReReach$(A, A, 0, \emptyset)$;

Procedure ReReach(A, A', n, S).

Input : $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$, $A, A' \in \mathsf{Args}$, $n \in \mathbb{N}_0$, $\mathsf{S} \subseteq \mathsf{Args} \times \mathsf{Args}$
Output: For each $B \in \mathsf{Args}$ with an attack path to A': $\mathsf{Reach}(A)$ and $\mathsf{Dist}(A, B)$
$\mathsf{Visited}_0 = \mathsf{S}$;
for $A^\star \in \mathit{Args}$ *s.t.* $(A^\star, A') \in \mathit{Att}$ *and* $(A^\star, A') \notin \mathit{Visited}$ **do**
  $\mathsf{Reach}(A) = \mathsf{Reach}(A) \cup \mathsf{Reach}(A^\star)$;
  $\mathsf{Dist}(A^\star, A) = \mathsf{Dist}(A^\star, A) \cup \{n + 1\}$;
  $\mathsf{Visited} = \mathsf{Visited} \cup \{(A^\star, A')\}$;
  ReReach$(A, A^\star, n + 1, \mathsf{Visited})$;
  $\mathsf{Visited} = \mathsf{Visited}_0$;

[9] Note that an attack-path is known as a *trail* in graph theory.

Algorithm 1 has some desirable properties. In particular, it is sound and complete (i.e., for the requested argument A it finds all the arguments from which A can be reached and the length of the attack-paths between those arguments), Theorem 1 and it runs in polynomial time, Theorem 2. The latter is useful since it shows that the computational complexity of computing the explanations for a certain semantics is not more complex than computing the acceptance of an argument and/or the extensions under that semantics [32]. See for the proofs of the results in this section Appendix A.2.

Theorem 1. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an argumentation framework. Then:*

1. *there is an attack-path from A to B of length n iff $n \in \mathsf{Dist}(A, B)$;*
2. *$A \in \mathsf{Reach}(B)$ iff there is an attack-path from A to B;*

By their definition, in order to request an explanation, the acceptability status of the argument or formula has to be known (i.e., whether it is skeptically or credulously (non-)acceptable). The presented algorithm does not compute this status. However, once it is known whether an explanation can be requested, the algorithm provides the necessary information to construct the basic explanations in polynomial time:

Theorem 2. *Algorithm 1 runs in polynomial time. In particular the time complexity is $\mathcal{O}(|Args| \cdot |Att|^2)$.*

Note that the run time of the algorithm is finite when Args and Att are finite. This is the case since for each $A \in \mathsf{Args}$ at most all attacks in Att are considered |Att| times.

Algorithm 1 determines for each argument the set of arguments from which it is reachable, as well as the distance between the arguments. From this we can define the notions from Definitions 5 and 7. We will denote by Reach_{odd} [resp. Reach_{even}] the arguments with odd [resp. even] distance to the considered argument (i.e., $\mathsf{Reach}_{odd}(A) = \{B \in \mathsf{Reach}(A) \mid \exists n \in \mathsf{Dist}(B, A) \text{ s.t. } n \text{ is odd}\}$ [resp. $\mathsf{Reach}_{even}(A) = \{B \in \mathsf{Reach}(A) \mid \exists n \in \mathsf{Dist}(B, A) \text{ s.t. } n \text{ is even}\}$]).

The next definition shows how Defending and NoDefAgainst can be defined in terms of the notions calculated by the algorithm.

Definition 10. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework, $A \in \mathsf{Args}$ and $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ an extension for some semantics Sem. Suppose that Algorithm 1 was run on $\mathcal{AF}$. Then:

- $\mathsf{Defending}(A) = \{B \in \mathsf{Reach}_{even}(A)\}$ denotes the set of arguments in Args that (in)directly defend A;
- $\mathsf{Defending}(A, \mathcal{E}) = \mathsf{Defending}(A) \cap \mathcal{E}$ denotes the set of arguments that (in)directly defend A in $\mathcal{E}$;
- $\mathsf{NoDefAgainst}(A, \mathcal{E}) = \{B \in \mathsf{Reach}_{odd}(A) \mid \mathcal{E} \cap \mathsf{Reach}_{odd}(B) = \emptyset\}$, denotes the set of all (in)direct attackers of A for which no defense exists from $\mathcal{E}$.

Example 13 *(Example 12 continued).* For the running example with the argumentation framework $\mathcal{AF}_1$, we have that $\mathsf{Reach}(A) = \{B, C, D, E, F\}$, with $\mathsf{Reach}_{even}(A) = \{C, E\}$ and $\mathsf{Reach}_{odd}(A) = \{B, D, F\}$. Moreover, $\mathsf{Dist}(C, A) = \{2, 4\}$ and $\mathsf{Dist}(E, B) = \{1, 3\}$. Indeed, $\mathsf{Defending}(A) = \{C, E\}$, $\mathsf{Defending}(A, \{A, C, F\}) = \{C\}$ and $\mathsf{NoDefAgainst}(A, \{B, D, F\}) = \{B, D, F\}$.

4. Varying the basic explanations

The basic explanations in the previous section were defined in terms of Defending and NoDefAgainst, from Definition 5 and 7. Intuitively, in the context of admissibility-based semantics, explanations based on these functions provide maximally relevant explanations for a (non-)accepted argument, that is, all the relevant attacking or defending arguments in the argumentation framework. However, depending on the domain and the specific arguments or types of argument, a user might prefer arguments that are directly relevant, that is, only direct attackers and defenders. For example, in $\mathcal{AF}_1$ (Fig. 1), we could answer the explanation-seeking question 'why is A not accepted?' by giving an explanation only in terms of a direct attacker: 'because there is no defense against attacker B'. Particularly in cases where there is a long chain of arguments that attack each other, considering only the direct attackers (or defenders) makes sense, as these will be most topically relevant to the explanandum.[10]

Another reason for varying what the relevant arguments in an explanation are, is that there are also semantics that are not based on the notion of defense like admissibility- or completeness-based semantics. For example, semantics based on naïve semantics from Definition 1 (e.g., stage, stage2 and CF2 semantics [5]) are about conflict rather than defense. So, for example, an explanation would then be 'B is not accepted because it attacks A and is attacked by C'. Furthermore, as we discussed in Remark 1, the notion of NoDefAgainst from Definition 7 does not work for all admissibility-based semantics.

In order to allow for different types of explanations, we need to be able to vary the functions Defending and NoDefAgainst in the definition of explanations (i.e., Definition 6 and 8). To this end the generic function $\mathbb{D}$ was employed in [16]. This function takes, like Defending and NoDefAgainst, as input an argument and a set of arguments and, in most cases, it returns a set of arguments. So for, for example, the credulous explanations we then have:

[10] This is similar to the idea of a direct cause in the literature on explanations and causation (e.g. [39,42]), where we might not always want to consider causes further down the causal chain of events for an explanation.

$\mathsf{SemAcc}^{\cup}(A,\mathcal{E}) = \mathbb{D}(A,\mathcal{E})$ for a given $A \in \mathsf{Args}$ and $\mathcal{E} \in \mathsf{SemWith}(A)$;

$\mathsf{SemNotAcc}^{\cup}(A) = \{\mathbb{D}(A,\mathcal{E}) \mid \mathcal{E} \in \mathsf{SemWithout}(A)\}$.

Where $\mathbb{D}$ can be Defending and NoDefAgainst, or some other function that returns the arguments that can be seen as a reason for the (non-)acceptance of A. In the rest of this section, we will propose some variations of $\mathbb{D}$ to (non-exhaustively) demonstrate the possibilities for the proposed framework. Note that, the variations discussed in this section are not the only possibilities. In Section 5 we will show how the content of explanations can be varied in ASPIC$^+$ and in Section 6 we introduce three notions for selecting the arguments in an explanation: minimality, necessity and sufficiency.

In examples, definitions, etc., where we treat acceptance and non-acceptance simultaneously, we will sometimes use the superscripts acc (for acceptance) and nacc (for non-acceptance) to specify the function $\mathbb{D}$. We do this to distinguish between the instantiations for acceptance and non-acceptance explanations.

4.1. Other notions of defense

In admissible and completeness-based semantics the notion of defense is fundamental in determining the arguments that can be accepted. The choice of $\mathbb{D}$ in the previous section (i.e., Defending and NoDefAgainst) was aimed at providing *all* relevant arguments for the (non-)acceptance of an argument, but there are also variations that still result in relevant sets of arguments, but that are not necessarily maximal. In order to keep the illustrations of variations of $\mathbb{D}$ simple we will provide small examples with abstract argumentation frameworks.

Acceptance and non-acceptance are mostly determined by direct conflicts. When an argument is defended against all its direct attackers, it is defended against all its attackers and can therefore be accepted. While, when an argument is not defended against a direct attacker it cannot be accepted given a completeness-based semantics. We introduce here variations to $\mathbb{D}$ that only consider the direct conflicts. In the context of acceptance, the explanations contain the direct defenders, while for non-acceptance such explanations result in sets of direct attackers to which the argument is not defended within some extension.

Definition 11. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework, $A \in \mathsf{Args}$ and $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ an extension for some semantics $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$. Then:

- $\mathsf{DirDefending}(A,\mathcal{E}) = \{B \in \mathcal{E} \mid B \text{ directly defends } A\}$ denotes the set of arguments that directly defend A in $\mathcal{E}$;
- $\mathsf{NoDirDefense}(A,\mathcal{E}) = \{B \in \mathsf{Args} \mid (B,A) \in \mathsf{Att} \text{ and } \nexists C \in \mathcal{E} \text{ such that } (C,B) \in \mathsf{Att}\}$ denotes the set of arguments that directly attack A and to which $\mathcal{E}$ has no defense.

Remark 4. Algorithm 1 from Section 3.4 can be applied to calculate explanations with $\mathbb{D} \in \{\mathsf{DirDefending}, \mathsf{NoDirDefense}\}$ as well:

- $\mathsf{DirDefending}(A,\mathcal{E}) = \{B \in \mathsf{Reach}_{\text{even}}(A) \mid 2 \in \mathsf{Dist}(B,A)\}$
- $\mathsf{NoDirDefense}(A,\mathcal{E}) = \{B \in \mathsf{Reach}_{\text{odd}}(A) \mid 1 \in \mathsf{Dist}(B,A) \text{ and } \nexists C \in \mathcal{E} \text{ such that } (C,B) \in \mathsf{Att}\}$.

By definition it follows that $\mathsf{DirDefending}(A,\mathcal{E}) \subseteq \mathsf{Defending}(A,\mathcal{E})$ and $\mathsf{NoDirDefense}(A,\mathcal{E}) \subseteq \mathsf{NoDefAgainst}(A,\mathcal{E})$.

Example 14 *(Examples 6 and 8 continued).* Recall that, for the argumentation framework $\mathcal{AF}_1$ from Example 1 and Fig. 1, where with $\mathbb{D}^{\mathsf{acc}} = \mathsf{Defending}$ and $\mathbb{D}^{\mathsf{nacc}} = \mathsf{NoDefAgainst}$ we had:

- $\mathsf{PrfAcc}^{\cup}(A,\mathcal{E}_1) = \{C,E\}$, $\mathsf{PrfAcc}^{\cup}(A,\mathcal{E}_2) = \{C\}$ and $\mathsf{PrfAcc}^{\cup}(A,\mathcal{E}_3) = \{E\}$ as acceptance explanations for A and $\mathsf{PrfAcc}^{\cup}(B,\mathcal{E}_4) = \{D,F\}$ as the acceptance explanation for B; and
- $\mathsf{PrfNotAcc}^{\cap}(A,\mathcal{E}_4) = \{B,D,F\}$ as a non-acceptance explanation for A.

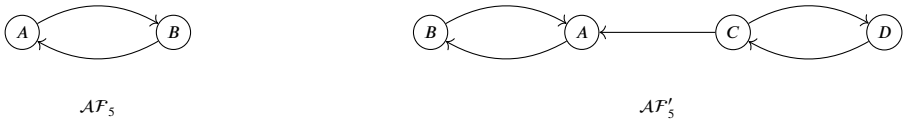
Now, when $\mathbb{D}^{\mathsf{acc}} = \mathsf{DirDefending}$ and $\mathbb{D}^{\mathsf{nacc}} = \mathsf{NoDirDefense}$:

- We still have that $\mathsf{PrfAcc}^{\cup}(A,\mathcal{E}_1) = \{C,E\}$, $\mathsf{PrfAcc}^{\cup}(A,\mathcal{E}_2) = \{C\}$ and $\mathsf{PrfAcc}^{\cup}(A,\mathcal{E}_3) = \{E\}$ and $\mathsf{PrfAcc}^{\cup}(B,\mathcal{E}_4) = \{D,F\}$, since both C and E are direct defenders of A and D and F are direct defenders of B;
- $\mathsf{PrfNotAcc}^{\cap}(A,\mathcal{E}_4) = \{B\}$, since out of $\{B,D,F\}$ only B directly attacks A.

Another way to consider defense is by only mentioning attackers against which the argument cannot defend itself, leading to the following variation of $\mathbb{D}$.

Definition 12. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework, $A, B \in \mathsf{Args}$ and let $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ be an extension for some semantics $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Sstb}, \mathsf{Egr}\}$. Then:

- $\mathsf{NoSelfDefense}(A,\mathcal{E}) = \{B \in \mathsf{NoDefAgainst}(A,\mathcal{E}) \mid A \text{ does not (in)directly attack } B\}$



$\mathcal{AF}_5$ $\mathcal{AF}'_5$

Fig. 5. Graphical representation of $\mathcal{AF}_5$ and $\mathcal{AF}'_5$.

denotes the set of arguments that attack A but for which no defense exists in $\mathcal{E}$ and that are not attacked by A itself.

Intuitively, a $\mathsf{NoSelfDefense}(A, \mathcal{E})$ explanations returns attackers of A to which there is no defense, and to defend A against these attacks other arguments than A itself are necessary.

Remark 5. Like for $\mathbb{D} \in \{\mathsf{Defending}, \mathsf{DirDefending}, \mathsf{NoDefAgainst}, \mathsf{NoDirDefense}\}$, Algorithm 1 can be applied to calculate explanations with $\mathbb{D} = \mathsf{NoSelfDefense}$:

- $\mathsf{NoSelfDefense}(A, \mathcal{E}) = \{B \in \mathsf{Reach}_{\mathrm{odd}}(A) \mid \mathcal{E} \cap \mathsf{Reach}_{\mathrm{odd}}(B) = \emptyset \text{ and } A \notin \mathsf{Reach}_{\mathrm{odd}}(B)\}$.

As was the case with NoDirDefense, we have that $\mathsf{NoSelfDefense}(A, \mathcal{E}) \subseteq \mathsf{NoDefAgainst}(A, \mathcal{E})$.

Example 15. Let $\mathcal{AF}_5 = \langle \mathsf{Args}_5, \mathsf{Att}_5\} \rangle$ with $\mathsf{Args}_5 = \{A, B\}$ and $\mathsf{Att}_5 = \{(A, B), (B, A)\}$ as in Fig. 5. Here $\mathsf{Prf}(\mathcal{AF}_5) = \{\mathcal{E}_1, \mathcal{E}_2\} = \{\{A\}, \{B\}\}$, $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}_2) = \{B\}$ for $\mathbb{D} \in \{\mathsf{NoDefAgainst}, \mathsf{NoDirDefense}\}$ since B (directly) attacks A but $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}_2) = \emptyset$ for $\mathbb{D} = \mathsf{NoSelfDefense}$ since by accepting A, A can indeed be concluded.

Now let $\mathcal{AF}'_5$ as in Fig. 5. Then $\mathsf{Prf}(\mathcal{AF}'_5) = \{\mathcal{E}'_1, \mathcal{E}'_2, \mathcal{E}'_3\} = \{\{A, D\}, \{B, C\}, \{B, D\}\}$, $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}'_2) = \{B, C\}$ and $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}'_3) = \{B\}$ for $\mathbb{D} \in \{\mathsf{NoDefAgainst}, \mathsf{NoDirDefense}\}$ since A is not defended against B in $\mathcal{E}'_3 = \{B, D\}$ and not defended against B and C in $\mathcal{E}'_2 = \{B, C\}$ and for $\mathbb{D} = \mathsf{NoSelfDefense}$ we have that $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}'_2) = \{C\}$, since in order to defend A, just accepting A is not enough, D is needed to defend against the attack from C.

4.2. Explanations under other semantics

The choice of the semantics is essential when reasoning with argumentation frameworks. Above we mainly focus on completeness-based semantics. However, even for some of the most applied completeness-based semantics (i.e., admissible and stable semantics) not all our results are applicable, as Remark 1 shows. We therefore propose here variations to NoDefAgainst that also see to the special cases of admissible and stable semantics. First, we introduce the following notation:

Notation 2. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, let $A \in \mathsf{Args}$ be an argument and $\mathsf{S} \subseteq \mathsf{Args}$ be a set of arguments. Then:

- $\mathsf{Relevant}(A, \mathsf{S}) = \{B \in \mathsf{S} \mid B \text{ is relevant for } A \text{ (Definition 4)}\}$.

With this notation the relevant arguments from a set of arguments can be determined.

Definition 13. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, let $A \in \mathsf{Args}$ be non-accepted and let $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for Sem as stated below. Then:

- $\mathsf{NoDefAgainstAdm}(A, \mathcal{E}) =$
$$\begin{cases} \mathsf{NoDefAgainst}(A, \mathcal{E}) & \text{if } \mathsf{NoDefAgainst}(A, \mathcal{E}) \neq \emptyset \\ \{\mathsf{Relevant}(A, \mathcal{E}' \setminus \mathcal{E}) \mid \mathcal{E}' \in \mathsf{AdmWith}(A) \text{ s.t. } \mathcal{E} \subset \mathcal{E}' \text{ and} & \text{otherwise} \\ \quad \mathsf{Relevant}(A, \mathcal{E}' \setminus \mathcal{E}) \neq \emptyset \text{ and } \nexists \mathcal{E}'' \in \mathsf{AdmWith}(A) \text{ s.t.} \\ \quad \mathcal{E} \subset \mathcal{E}'' \subset \mathcal{E}' \text{ and } \mathsf{Relevant}(A, \mathcal{E}'' \setminus \mathcal{E}) \neq \emptyset\} \end{cases}$$
- $\mathsf{NoDefAgainstStb}(A, \mathcal{E}) = \mathsf{NoDefAgainst}(A, \mathcal{E}) \cup \{\mathsf{S} \subseteq \mathsf{Args} \mid$ there is an attack-path from A to A of odd length along the arguments in S and $\mathcal{E}$ does not attack $\mathsf{S}\}$.

For admissibility, when NoDefAgainst is not empty, it contains a reason for the non-acceptance of A, but if NoDefAgainst is empty, A might still be accepted given $\mathcal{E}$ (i.e., $\mathcal{E}$ might not be complete). This is formalized in the second case. In particular, when the second case is applied, $\mathsf{NoDefAgainstAdm}(A, \mathcal{E})$ contains the relevant arguments for A that are part of a minimal but strict superset of $\mathcal{E}$. This way, explanations based on this variation will collect the arguments that influence the acceptance of A, even though these are not part of the current admissible extension. In contrast, for stable semantics, the existence of an odd cycle can provide additional information on the non-acceptance of argument rather than a different reason for the non-acceptance. Therefore, an odd cycle containing A and no arguments attacked by $\mathcal{E}$ is added.

Remark 6. Calculating explanations with Algorithm 1 from Section 3.4 for the variations $\mathbb{D} \in \{\mathsf{NoDefAgainstAdm}, \mathsf{NoDefAgainstStb}\}$ is less straightforward than it was for the variations in the previous section. The reason for this is that the above definition requires additional notions (e.g., Relevant and the collection of arguments along an attack-path). We therefore do not provide definitions and leave a precise calculation up to the reader.

Example 16. In Remark 1 we had an argumentation framework with arguments A and B and no attacks. Then $\mathsf{NoDefAgainst}(A, \{B\}) = \emptyset$. With the new variation of $\mathbb{D}$ we have now: $\mathsf{NoDefAgainstAdm}(A, \{B\}) = \{A\}$, since $\{A, B\}$ is also admissible.

In the same remark, the attack (B, B) was added. While we had that $\mathsf{NoDefAgainst}(B, \{A\}) = \emptyset$, with the new definition $\mathsf{NoDefAgainstStb}(B, \{A\}) = \{B\}$, since B attacks itself.

Up until now the variations to $\mathbb{D}$ were based on the notion of defense which is particularly important for semantics that are based on admissibility and completeness. However, there are other types of semantics as well. An important family of semantics is based on naïve semantics (recall Definition 1), which are not about defense but rather about conflict-freeness. In addition to naïve semantics, other semantics based on conflict-freeness (that result in specific naïve extensions) are stage, stage2 and CF2 semantics [5].

Since naïve semantics is not about defended arguments but only about conflict-free sets of arguments, $\mathbb{D}$ can be defined in terms of conflict between argument. For this consider the following definition:

Definition 14. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, $A \in \mathsf{Args}$ and $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ an extension for some $\mathsf{Sem} \in \{\mathsf{Nav}, \mathsf{Stg}\}$. Then:

- $\mathsf{ExtAttacking}(A, \mathcal{E}) = \{B \in \mathcal{E} \mid (A, B) \in \mathsf{Att} \text{ or } (B, A) \in \mathsf{Att}\}$ denotes the set of arguments from $\mathcal{E}$ that attack A or are attacked by A.

This set can be varied by selecting only the attacking or only the attacked arguments:

- $\mathsf{ExtAttacking}^{+}(A, \mathcal{E}) = \{B \in \mathsf{ExtAttacking}(A, \mathcal{E}) \mid (A, B) \in \mathsf{Att}\}$ and $\mathsf{ExtAttacking}^{-}(A, \mathcal{E}) = \{B \in \mathsf{ExtAttacking}(A, \mathcal{E}) \mid (B, A) \in \mathsf{Att}\}$ denote the set of arguments from $\mathcal{E}$ that are directly attacked by resp. that directly attack A.

Remark 7. Given $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ and $A \in \mathsf{Args}$, let $\mathsf{Sem} = \mathsf{Nav}$. By definition of naïeve semantics, A is credulously accepted if it does not attack itself and skeptically accepted if it is not attacked at all. There are therefore no relevant arguments for the credulous/skeptical acceptance of argument A, and hence no explanations in terms of relevant arguments. We only consider non-acceptance here, since these explanations do contain relevant arguments: those arguments that A has a conflict with.

Remark 8. Although Algorithm 1 can be applied to calculate ExtAttacking (i.e., by determining that $A \in \mathsf{Reach}_{\mathsf{odd}}(B)$ and $1 \in \mathsf{Dist}(A, B)$ or vice versa), it is more efficient to search the attack relation of the argumentation framework.

Example 17. Consider $\mathsf{Sem} = \mathsf{Nav}$ and $\mathbb{D} = \mathsf{ExtAttacking}$. Note that, intuitively, an argument A is not skeptically accepted if there is some conflict-free set that would no longer be conflict-free if A was added. For the argumentation framework $\mathcal{AF}_1$ from Example 1 we have that:

- $\mathsf{NavNotAcc}^{\cap}(B, \mathcal{E}_1) = \{A, C, E\}$, $\mathsf{NavNotAcc}^{\cap}(B, \mathcal{E}_2) = \{A, C\}$ and $\mathsf{NavNotAcc}^{\cap}(B, \mathcal{E}_3) = \{A, E\}$. Moreover, we also have that $\mathsf{NavNotAcc}^{\cap}(B, \mathcal{E}_1) = \{C, E\}$, $\mathsf{NavNotAcc}^{\cap}(B, \mathcal{E}_2) = \{C\}$ and $\mathsf{NavNotAcc}^{\cap}(B, \mathcal{E}_3) = \{E\}$ for $\mathbb{D} = \mathsf{ExtAttacking}^{-}$ and $\mathsf{NavNotAcc}^{\cap}(B, \mathcal{E}) = \{A\}$ for $\mathcal{E} \in \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3\}$ and $\mathbb{D} = \mathsf{ExtAttacking}^{+}$. Thus, the argument B is not skeptically accepted, because there are extensions that contain A, C and/or E, which attack or are attacked by B.

For a non-accepted argument A, explanations with $\mathbb{D} = \mathsf{ExtAttacking}$ provide explanations in terms of the conflicts A has in a specific extension.

4.3. Some results for the variations of $\mathbb{D}$

In Section 3.3 some properties of the explanations with $\mathbb{D}^{\mathsf{acc}} = \mathsf{Defending}$ and $\mathbb{D}^{\mathsf{nacc}} = \mathsf{NoDefAgainst}$ were shown. In this section we further investigate some of those properties for the variations of $\mathbb{D}$. The proofs of the results in this section are provided in Appendix A.3.

As was mentioned after their definition, the variations of $\mathbb{D}$ based on the notion of defense are closely related to Defending and NoDefAgainst:

Remark 9. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, let $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for some $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and let $A \in \mathsf{Args}$. Then:

- $\mathsf{DirDefending}(A, \mathcal{E}) \subseteq \mathsf{Defending}(A, \mathcal{E})$;

- $\mathsf{NoDirDefense}(A, \mathcal{E}) \subseteq \mathsf{NoDefAgainst}(A, \mathcal{E})$ and, similarly $\mathsf{NoSelfDefense}(A, \mathcal{E}) \subseteq \mathsf{NoDefAgainst}(A, \mathcal{E})$.

Note however that NoDirDefense and NoSelfDefense are not related, it might be the case that: $\mathsf{NoDirDefense}(A, \mathcal{E}) \cap \mathsf{NoSelfDefense}(A, \mathcal{E}) = \emptyset$ (recall Example 15).

The next corollary (as well as Corollary 3 below) shows that acceptance [resp. non-acceptance] explanations are still well-behaved when the variations of $\mathbb{D}$ introduced in this section are applied instead of Defending [resp. NoDefAgainst]. Thus, like for the basic explanations, explanations under one semantics are related to explanations under another semantics in the same way as extensions under the first semantics are related to extensions under the other semantics.

Corollary 1. *Let $\mathcal{AF} = \langle Args, Att \rangle$ be an argumentation framework, $A \in Args$ and let $\mathbb{D} = \mathsf{DirDefending}$. Then:*

1. $\mathsf{GrdAcc}(A) \subseteq \bigcap \mathsf{SemAcc}^{\cap}(A)$ *for* $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$.
2. $\mathsf{StbAcc}^{\cap}(A) \subseteq \mathsf{SstbAcc}^{\cap}(A) \subseteq \mathsf{PrfAcc}^{\cap}(A) \subseteq \mathsf{CmpAcc}^{\cap}(A)$.
3. *For each* $\mathsf{S} \in \mathsf{CmpAcc}^{\cap}(A)$ *there is an* $\mathsf{S}' \in \mathsf{PrfAcc}^{\cap}(A)$ *such that* $\mathsf{S} \subseteq \mathsf{S}'$.

A counterpart of Proposition 2 does not hold for $\mathbb{D} = \mathsf{DirDefending}$. This is the case because direct defense might vary from argument to argument, even if the arguments are part of the same attack path(s). However, we do have that acceptance explanations for $\mathbb{D} = \mathsf{DirDefending}$ are only empty when the argument is not attacked at all.

Corollary 2. *Let $\mathcal{AF} = \langle Args, Att \rangle$ be an AF, let $A \in Args$ be such that A is accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ *and* $\star \in \{\cap, \cup\}$ *and let* $\mathbb{D} = \mathsf{DirDefending}$. *Then* $\mathsf{SemAcc}^{\cap}(A) = \{\emptyset\}$ *and* $\mathsf{SemAcc}^{\cup}(A, \mathcal{E}) = \emptyset$ *for any* $\mathcal{E} \in \mathsf{SemWith}(A)$ *iff there is no* $B \in Args$ *such that* $(B, A) \in Att$.

The next corollaries are on non-acceptance explanations for the variations $\mathbb{D} \in \{\mathsf{NoDirDefense}, \mathsf{NoSelfDefense}\}$ and the counterparts of Proposition 4 and 5. First, it is shown that the non-acceptance explanations with $\mathbb{D} \in \{\mathsf{NoDirDefense}, \mathsf{NoSelfDefense}\}$ also behave predictable with respect to the chosen semantics.

Corollary 3. *Let $\mathcal{AF} = \langle Args, Att \rangle$ be an argumentation framework, $A \in Args$ be non-accepted, $\star \in \{\cap, \cup\}$ and let $\mathbb{D} \in \{$*NoDirDefense, NoSelfDefense*$\}$. Then:*

- $\mathsf{GrdNotAcc}(A) \subseteq \bigcap \mathsf{CmpNotAcc}^{\cup}(A)$.
- $\mathsf{SstbNotAcc}^{\cup}(A) \subseteq \mathsf{PrfNotAcc}^{\cup}(A) \subseteq \mathsf{CmpNotAcc}^{\cup}(A)$.

Unlike non-acceptance explanations with $\mathbb{D} \in \{\mathsf{NoDefAgainst}, \mathsf{NoDirDefense}\}$, a non-acceptance explanation for $\mathbb{D} = \mathsf{NoSelfDefense}$ can be empty. This is the case since an argument might not be part of some extension even though it defends itself against all its attackers. Like for empty acceptance explanations, the implementation of these explanations into an application needs to be extended to provide explanations when the explanation is empty.

Corollary 4. *Let $\mathcal{AF} = \langle Args, Att \rangle$ be an argumentation framework and let $A \in Args$ be such that A is non-accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ *and* $\star \in \{\cap, \cup\}$. *Then:*

- *if* $\mathbb{D} = \mathsf{NoDirDefense}$ *then* $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) \neq \emptyset$ *for any extension* $\mathcal{E} \in \mathsf{SemWithout}(A)$ *and* $\mathsf{SemNotAcc}^{\cup}(A) \neq \{\emptyset\}$*; and*
- *if* $\mathbb{D} = \mathsf{NoSelfDefense}$ *then* $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) = \emptyset$ *for any extension* $\mathcal{E} \in \mathsf{SemWithout}(A)$ *or* $\mathsf{SemNotAcc}^{\cup}(A) = \{\emptyset\}$ *implies that for all* $B \in Args$ *such that* $(B, A) \in Att$, $(A, B) \in Att$ *as well.*

5. Explanations in structured argumentation settings

In the previous sections we discussed explanations in terms of abstract argumentation frameworks. However, it has been argued that we should take the structure of the arguments and the nature of attacks into account, see, e.g., [2,20,48,57,58]. Most of the ideas on explanations for abstract argumentation can be applied to any (structured) approach to argumentation that results in an abstract argumentation framework. In this section we make this explicit. Moreover, we show that the explanations can be refined when the structure of the arguments is taken into account.

To show the advantages of explanations for structured argumentation, we choose ASPIC$^+$ [56], which allows for two types of premises – *axioms* that cannot be questioned and *ordinary premises* that can be questioned – and two types of rules – *strict* rules that cannot be questioned and *defeasible rules*. We choose ASPIC$^+$ as the structured argumentation approach in this paper since, due to the variety of input it can be given, it allows to vary the form of the explanations in many ways (see Section 5.3). After recalling the most important notions from ASPIC$^+$ (Section 5.1), we discuss how the basic explanations can be applied to the structured setting (Section 5.2). We then show how these basic explanations can be varied (Section 5.3).

5.1. ASPIC$^+$

ASPIC$^+$ has extensively been studied in the literature, see e.g., [50,51] for surveys. Most of the definitions in this section are based on [49,56]. An ASPIC$^+$ setting starts from an *argumentation system* (AS). Arguments are then constructed in an argumentation system from a knowledge based, which together form the *argumentation theory*. Based on the structure of the constructed arguments, the attack relation is determined, with which argumentation frameworks can be defined.

Definition 15. An *argumentation system* is a tuple $\mathrm{AS} = \left\langle \mathcal{L}, \bar{}, \mathcal{R}, n \right\rangle$, where:

- $\mathcal{L}$ is a propositional language;
- $\bar{}$ is a function that assigns to each formula in $\mathcal{L}$ a set of formulas such that, for $\phi, \psi \in \mathcal{L}$:
  - ϕ is a *contrary* of ψ if $\phi \in \overline{\psi}$ but $\psi \notin \overline{\phi}$;
  - ϕ is a *contradictory* of ψ if $\phi \in \overline{\psi}$ and $\psi \in \overline{\phi}$, this will be denoted by $\phi = -\psi$;
  - each $\phi \in \mathcal{L}$ is assumed to have at least one contradictory.
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\phi_1, \ldots, \phi_n \rightarrow \phi$ resp. $\phi_1, \ldots, \phi_n \Rightarrow \phi$, such that $\{\phi_1, \ldots, \phi_n, \phi\} \subseteq \mathcal{L}$ and $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
  Where $r \in \mathcal{R}$, we denote by:
  - $\mathsf{Ant}(r) = \{\phi_1, \ldots, \phi_n\}$ the *antecedents* of the rule;
  - $\mathsf{Cons}(r) = \phi$ the *consequent* of the rule; and
  - $\mathsf{Rules}(\mathcal{R}, \phi) = \{r \in \mathcal{R} \mid \mathsf{Cons}(r) = \phi\}$ the set of rules with ϕ as consequent.
- $n : \mathcal{R}_d \rightarrow \mathcal{L}$ is a naming convention for defeasible rules, where $n(r)$ is a well-formed formula in $\mathcal{L}$ which says that $r \in \mathcal{R}_d$ is applicable.

A *knowledge base* in an argumentation system $\left\langle \mathcal{L}, \bar{}, \mathcal{R}, n \right\rangle$ is a set of formulas $\mathcal{K} \subseteq \mathcal{L}$ which contains two disjoint subsets: $\mathcal{K} = \mathcal{K}_p \cup \mathcal{K}_n$, the set of *axioms* $\mathcal{K}_n$ and the set of *ordinary premises* $\mathcal{K}_p$.

The combination of an argumentation system and a knowledge base is called an argumentation theory:

Definition 16. An *argumentation theory* is a pair $\mathrm{AT} = \langle \mathrm{AS}, \mathcal{K} \rangle$, where AS is an argumentation system and $\mathcal{K}$ is a knowledge base.

Arguments in ASPIC$^+$ are then defined relative to an argumentation theory $\mathrm{AT} = \langle \mathrm{AS}, \mathcal{K} \rangle$.

Definition 17. An *argument* A on the basis of a knowledge base $\mathcal{K}$ in an argumentation system $\left\langle \mathcal{L}, \bar{}, \mathcal{R}, n \right\rangle$ is:

1. ϕ if $\phi \in \mathcal{K}$, where $\mathsf{Prem}(A) = \mathsf{Sub}(A) = \{\phi\}$, $\mathsf{Conc}(A) = \phi$, $\mathsf{Rules}(a) = \emptyset$ and $\mathsf{TopRule}(A) =$ undefined;
2. $A_1, \ldots, A_n \rightsquigarrow \psi$, where $\rightsquigarrow \in \{\rightarrow, \Rightarrow\}$, if $A_1, \ldots, A_n$ are arguments such that there exists a rule $\mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \rightsquigarrow \psi$ in $\mathcal{R}_s$ if $\rightsquigarrow = \rightarrow$ and in $\mathcal{R}_d$ if $\rightsquigarrow = \Rightarrow$. We denote by:
   - $\mathsf{Prem}(A) = \mathsf{Prem}(A_1) \cup \ldots \cup \mathsf{Prem}(A_n)$ the set of *premises* of A;
   - $\mathsf{Conc}(A) = \psi$ the *conclusion* of A;
   - $\mathsf{Sub}(A) = \mathsf{Sub}(A_1) \cup \ldots \cup \mathsf{Sub}(A_n) \cup \{A\}$ the set of *subarguments* of A;
   - $\mathsf{Rules}(A) = \mathsf{Rules}(A_1) \cup \ldots \cup \mathsf{Rules}(A_n) \cup \{\mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \rightsquigarrow \psi\}$ the set of *rules* applied in the construction of A;
   - $\mathsf{DefRules}(A) = \{r \in \mathcal{R}_d \mid r \in \mathsf{Rules}(A)\}$ the set of *defeasible rules* applied in the construction of A; and
   - $\mathsf{TopRule}(A) = \mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \rightsquigarrow \psi$ the *top rule* (i.e., the last rule applied in the construction) of A.

The above notations can be generalized to sets. For example, given a set S of arguments $\mathsf{Prem}(\mathsf{S}) = \bigcup \{\mathsf{Prem}(A) \mid A \in \mathsf{S}\}$, $\mathsf{Conc}(\mathsf{S}) = \{\mathsf{Conc}(A) \mid A \in \mathsf{S}\}$ and $\mathsf{DefRules}(\mathsf{S}) = \bigcup \{\mathsf{DefRules}(A) \mid A \in \mathsf{S}\}$.

Example 18. Let $\mathrm{AS}_1 = \langle \mathcal{L}_1, \neg, \mathcal{R}_1, n \rangle$ be an argumentation system where $\neg$ denotes classical negation and where $\mathcal{R}_1$ contains six defeasible rules such that, with $\mathcal{K}_1 = \mathcal{K}_p^1 = \{r, s, t, v, w\}$ the following arguments can be derived:

$$A : w \overset{d_1}{\Rightarrow} u \qquad B : D, F \overset{d_2}{\Rightarrow} \neg n(d_1) \qquad C : r, s \overset{d_3}{\Rightarrow} q$$

$$D : v \overset{d_4}{\Rightarrow} \neg q \qquad E : r, t \overset{d_5}{\Rightarrow} \neg p \qquad F : v \overset{d_6}{\Rightarrow} p.$$

The arguments based on the knowledge base only (i.e., without the application of any rule) are not mentioned here. This does not cause problems since, as will become clear after the next definition, these arguments do not attack any argument, nor are they attacked by another argument.

Attacks on an argument are based on the rules and premises applied in the construction of that argument.

Definition 18. Let A and B be two arguments. Then A *attacks* an argument B iff A *undercuts*, *rebuts* or *undermines* B, where:

- A *undercuts* B (on B') iff $\mathsf{Conc}(A) = \overline{n(r)}$ for some $B' \in \mathsf{Sub}(B)$ such that B''s top rule r is defeasible;
- A *rebuts* B (on B') iff $\mathsf{Conc}(A) = \overline{\phi}$ for some $B' \in \mathsf{Sub}(B)$ of the form $B'_1, \ldots, B'_n \Rightarrow \phi$ and A *contrary-rebuts* B iff $\mathsf{Conc}(A)$ is a contrary of ϕ;
- A *undermines* B (on ϕ) iff $\mathsf{Conc}(A) = \overline{\phi}$ for some $\phi \in \mathsf{Prem}(B) \setminus \mathcal{K}_n$ and A *contrary-undermines* B iff $\mathsf{Conc}(A)$ is a contrary of ϕ.

Intuitively, when A undercuts B, it attacks B in the application of a defeasible rule used in the construction of B; when A rebuts B, it attacks B in (one of) its (sub)conclusions; and when A undermines B it attacks B in one of the premises used in the construction of B.

Example 19. For the arguments from Example 18, derived from AS_1 and $\mathcal{K}_1$, we have that:

- B undercuts A,
- C and D as well as E and F rebut each other.

Note that, because of our choice of contrariness relation (i.e., classical negation), the attacks are not contrary-rebuts or contrary-undermines, since classical negation always results in contradictories.

Often, ASPIC$^+$ comes equipped with a preference ordering (see, e.g., [49,56]). Since our explanations are based on the argumentation framework constructed from an ASPIC$^+$ setting, the explanations do not depend on a possible preference relation. For the sake of conciseness we will therefore not introduce a preference relation here. The correspondence between argumentation theories and Dung-style argumentation frameworks can then be defined as follows.

Definition 19. Given an argumentation theory $\mathrm{AT} = \langle \mathrm{AS}, \mathcal{K} \rangle$, the corresponding Dung-style abstract argumentation framework is $\mathcal{AF}(\mathrm{AT}) = \langle \mathsf{Args}, \mathsf{Att} \rangle$, where Args is the set of arguments that can be constructed from AT, $(A, B) \in \mathsf{Att}$ iff $A, B \in \mathsf{Args}$ and A attacks B as defined in Definition 18.

Example 20. From the argumentation system AS_1 and the knowledge base $\mathcal{K}_1$ from Example 18 we obtain the argumentation theory $\mathrm{AT}_1 = \langle \mathrm{AS}_1, \mathcal{K}_1 \rangle$. Note that we have that $\mathcal{AF}(\mathrm{AT}_1) = \langle \mathsf{Args}_1, \mathcal{A}_1 \rangle$ as in Example 1 and that Fig. 1 represents this argumentation framework (again ignoring the arguments based on the knowledge base).

Dung-style semantics, as in Definition 1, can be applied in the same way as they are applied to abstract argumentation frameworks. Recall Example 2 for a discussion. Note that, unlike the abstract framework, in the structured framework the grounded extension is not empty: the arguments based on the knowledge base that we ignored so far are part of the grounded, as well as any other completeness-based extension.

Like for abstract argumentation it can be determined which arguments are (not) skeptically or credulously accepted. Moreover, the notions of (in)direct attack and defense (Definition 3) as well as relevance (Definition 4) can be applied to argumentation frameworks derived from argumentation theories as well. In addition, acceptance and non-acceptance can be defined for formulas:

Definition 20. Let $\mathcal{AF}(\mathrm{AT}) = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, based on some argumentation theory AT, let Sem be a semantics and suppose that $\mathsf{Sem}(\mathcal{AF}) \neq \emptyset$ and let $\phi \in \mathcal{L}$. Then ϕ is:

- *Credulously accepted*: iff $\phi \in \bigcup \mathsf{Concs}(\mathsf{Sem}(\mathcal{AF}(\mathrm{AT})))$, that is: there is some Sem-extension with an argument A such that $\mathsf{Conc}(A) = \phi$;
- *Skeptically accepted*: iff $\phi \in \bigcap \mathsf{Concs}(\mathsf{Sem}(\mathcal{AF}(\mathrm{AT})))$, that is: in each Sem-extension there is an argument with conclusion ϕ;
- *Not credulously accepted*: iff $\phi \notin \bigcup \mathsf{Concs}(\mathsf{Sem}(\mathcal{AF}(\mathrm{AT})))$, that is: there is no Sem-extension with an argument with conclusion ϕ;
- *Not skeptically accepted*: iff $\phi \notin \bigcap \mathsf{Concs}(\mathsf{Sem}(\mathcal{AF}(\mathrm{AT})))$, that is: there is some Sem-extension without an argument with conclusion ϕ.

When arbitrary or clear from the context (e.g., when $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Idl}, \mathsf{Egr}\}$), we will write that the formula is (non-)accepted. Otherwise, like before, we will indicate (not) skeptically [resp. credulously] accepted by $\cap$ [resp. $\cup$].

Example 21. For the argumentation framework $\mathcal{AF}(\mathrm{AT}_1)$ from Example 20, with the arguments from Example 18, we have that:

1. $\phi \in \{r, s, t, v, w\}$ are skeptically accepted for $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$; while
2. $\phi \in \{p, \neg p, q, \neg q, u\}$ are credulously accepted but not skeptically accepted for $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Stb}, \mathsf{Sstb}\}$; and
3. there is no formula that is not credulously accepted.

This follows since the arguments based on $\mathcal{K}_1$ (i.e., those not mentioned in Example 18) are part of every completeness-based extension and all other arguments are part of at least one extension but not of all extensions (for $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Stb}, \mathsf{Sstb}\}$).

5.2. Basic explanations for formulas

While Dung-style argumentation frameworks can be derived from argumentation theories (recall Definition 16), the acceptance of arguments and formulas might differ. For example, there might be several arguments for a formula ϕ and although none of these arguments is skeptically accepted, ϕ might be skeptically accepted because in each extension there is at least one argument for ϕ (Definition 20). Similarly, explanations for the (non-)acceptance of arguments and formulas can differ. In order to define these explanations, we introduce the following notation.

Notation 3. Let $\mathcal{AF}(\mathrm{AT}) = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, $\phi \in \mathcal{L}$ and let $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Grd}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$. Then:

- $\mathsf{AllArgs}(\phi) = \{A \in \mathsf{Args} \mid \mathsf{Conc}(A) = \phi\}$ denotes the set of all arguments of $\mathcal{AF}(\mathrm{AT})$ with conclusion ϕ;
- $\mathsf{SemAccept}(\phi) = \mathsf{AllArgs}(\phi) \cap \bigcup \mathsf{Sem}(\mathcal{AF}(AT))$ denotes the set of all arguments of $\mathcal{AF}(\mathrm{AT})$ with conclusion ϕ that are part of at least one Sem-extension (i.e., that are credulously accepted);
- $\mathsf{SemWith}(\phi) = \bigcup\{\mathsf{SemWith}(A) \mid A \in \mathsf{AllArgs}(\phi)\}$ denotes the set of Sem-extensions that contain an argument with conclusion ϕ; and
- $\mathsf{SemWithout}(\phi) = \bigcap\{\mathsf{SemWithout}(A) \mid A \in \mathsf{AllArgs}(\phi)\}$ denotes the set of Sem-extensions that do not contain any argument with conclusion ϕ.

Example 22. For the argumentation framework $\mathcal{AF}(\mathrm{AT}_1)$ with the arguments from Example 18 and $\mathsf{Sem}(\mathcal{AF}(\mathrm{AT}_1)) = \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4\} = \{\{A, C, E\}, \{A, C, F\}, \{A, D, E\}, \{B, D, F\}\}$, we have that: $\mathsf{AllArgs}(p) = \{F\}$, $\mathsf{SemWith}(p) = \{\{A, C, F\}, \{B, D, F\}\}$ and $\mathsf{SemWithout}(p) = \{\{A, C, E\}, \{A, D, E\}\}$ for $\mathsf{Sem} \in \{\mathsf{Prf}, \mathsf{Stb}, \mathsf{Sstb}\}$.

One important difference in formula explanations is that for a formula ϕ it makes sense to also consider *existence explanations*. After all, an intuitive answer to the question 'why is ϕ accepted?' is to provide (all the) arguments that have ϕ as their conclusion as reasons for why the formula can be derived in the first place.

Definition 21 *(Existence explanation).* Let $\mathcal{AF}(\mathrm{AT}) = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework based on the argumentation theory AT and let $\phi \in \mathcal{L}$ be such that ϕ is $\star$-accepted for $\mathsf{Sem} \in \{\mathsf{Nav}, \mathsf{Stg}, \mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$.

$$\mathsf{SemExist}^{\star}(\phi) = \mathsf{SemAccept}(\phi).$$

For the considered formula ϕ, these existence explanations collect all the arguments for ϕ that are part of some Sem-extension. It is important to note that the definition adds an essential condition to the notation SemAccept: ϕ should be accepted w.r.t. Sem and $\star$. Thus, $\mathsf{SemExist}^{\cup}(\phi)$ returns all the arguments for ϕ that are part of some Sem-extension and could therefore be reasons for the credulous acceptance of ϕ. $\mathsf{SemExist}^{\cap}(\phi)$ returns for each Sem-extension at least one arguments for ϕ (since ϕ is skeptically accepted).

Example 23. For the argumentation framework $\mathcal{AF}(\mathrm{AT}_1)$, we have that only the formulas in $\mathcal{K}_1$ are skeptically accepted. Additionally, we have explanations for why, for example, u and $\neg n(d_1)$ are credulously accepted under Prf:

- $\mathsf{PrfExist}^{\cup}(u) = \{A\}$ since A is the only argument for u in Args_1;
- $\mathsf{PrfExist}^{\cup}(\neg n(d_1)) = \{B\}$ since B is the only argument for $\neg n(d_1)$ in Args_1.

5.2.1. Explanations for accepted formulas

Formula explanations are very similar to the basic argument explanations introduced in Section 3, but in addition to accounting for all the extensions, now the possible existence of several arguments for one formula has to be taken into account as well. Thus, the question why some formula ϕ is accepted is then answered by an explanation of the form 'because argument A has ϕ as its conclusion, and A is accepted because it is defended against its attackers by arguments $B_1, \ldots, B_n$'. The definitions in this section use the generic function $\mathbb{D}$ as introduced in Section 4, in the examples we will take $\mathbb{D}^{\mathsf{acc}} = \mathsf{Defending}$ and $\mathbb{D}^{\mathsf{nacc}} = \mathsf{NoDefAgainst}$. In the next section we will further extend this definition by allowing to vary the form of the explanation.

Definition 22 *(Formula acceptance explanation).* Let $\mathcal{AF}(\mathrm{AT}) = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework based on the argumentation theory AT, let $\phi \in \mathcal{L}$ be such that ϕ is $\star$-accepted and $\mathcal{E}_\phi \in \mathsf{Sem}(\mathcal{AF})$ such that $\phi \in \mathsf{Concs}(\mathcal{E}_\phi)$, for $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$.

$$\mathsf{SemAcc}^{\cap}(\phi) = \{\langle A, \mathbb{D}(A, \mathcal{E})\rangle \mid A \in \mathsf{SemExist}^{\cap}(\phi) \text{ and } \mathcal{E} \in \mathsf{SemWith}(A)\}$$

$$\mathsf{SemAcc}^{\cup}(\phi, \mathcal{E}_\phi) = \{\left\langle A, \mathbb{D}(A, \mathcal{E}_\phi)\right\rangle \mid A \in \mathcal{E}_\phi \text{ and } \mathsf{Conc}(A) = \phi\}.$$

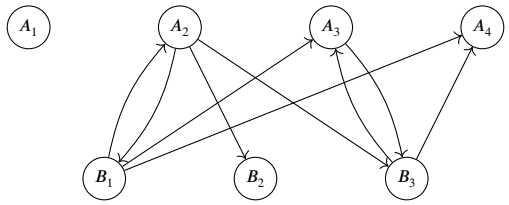


Fig. 6. Graphical representation of the argumentation framework $\mathcal{AF}(\mathrm{AT}_6)$ from Example 25.

For the considered formula ϕ, the (skeptical) $\cap$-explanation collects for each accepted argument A with conclusion ϕ and each Sem-extension $\mathcal{E}$ that contains A the set $\mathbb{D}(A,\mathcal{E})$ and provides this as a pair with the argument A (why is ϕ skeptically accepted under Sem or, in other words, which arguments for ϕ exist and which arguments defend these arguments in each extension?). Compared to Definition 6 for argument explanations, the main difference is that here explanations are pairs and all accepted arguments for ϕ are part of the explanation. As a result it explains the existence of an argument for ϕ and the acceptance of all accepted arguments for ϕ for all extensions in which the arguments are accepted. The (credulous) $\cup$-explanation returns for each argument A in the given extension $\mathcal{E}_\phi$ with conclusion ϕ the pair containing A and $\mathbb{D}(A,\mathcal{E}_\phi)$ (why is ϕ credulously accepted under Sem or, in other words, which arguments for ϕ exist in the given extension $\mathcal{E}_\phi$ and which arguments from $\mathcal{E}_\phi$ defend these arguments?). It thus explains the existence of arguments for ϕ and why these arguments are accepted in the given Sem-extension.

Example 24. Recall from Example 21 that, for the argumentation framework $\mathcal{AF}(\mathrm{AT}_1)$ from Example 20, none of the arguments from Example 18 were skeptically accepted, but all are credulously accepted. We therefore have, for $\mathsf{Sem} \in \{\mathsf{Prf}, \mathsf{Stb}, \mathsf{Sstb}\}$:

- $\mathsf{SemAcc}^{\cup}(u,\mathcal{E}_1) = \{\langle A, \{C,E\}\rangle\}$, $\mathsf{SemAcc}^{\cup}(u,\mathcal{E}_2) = \{\langle A, \{C\}\rangle\}$ and $\mathsf{SemAcc}^{\cup}(u,\mathcal{E}_3) = \{\langle A, \{E\}\}$ since A is the only argument for u and the Sem-extensions with A are $\mathcal{E}_1 = \{A,C,E\}$, $\mathcal{E}_2 = \{A,C,F\}$ and $\mathcal{E}_3 = \{A,D,E\}$;
- $\mathsf{SemAcc}^{\cup}(\neg n(d_1),\mathcal{E}_4) = \{\langle B, \{D,F\}\rangle\}$ since B is the only argument for $\neg n(d_1)$ and the only Sem-extension with B is $\mathcal{E}_4 = \{B,D,F\}$.

So u is accepted because of the existence of argument A, which is defended by C and E, C or E in the extensions $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_3$ respectively. Similarly, $\neg n(d_1)$ is accepted because of the existence of argument B, which is defended by D and F in the extension $\mathcal{E}_4$.

In order to show explanations when there are more arguments for one conclusion, we recall the following example, from [16, Example 3].

Example 25. Let $\mathrm{AS}_6 = \langle \mathcal{L}_6, \neg, \mathcal{R}_6, n\rangle$ be an argumentation system where $\neg$ denotes classical negation and where $\mathcal{R}_6$ contains five defeasible rules such that, with $\mathcal{K}_6 = \mathcal{K}_n^6 \cup \mathcal{K}_p^6$ where $\mathcal{K}_n^6 = \{t\}$ and $\mathcal{K}_p^6 = \{r\}$ the following arguments can be derived:

$$A_1 : t \qquad A_2 : A_1 \overset{d_3}{\Rightarrow} \neg r \qquad A_3 : A_1, A_2 \overset{d_4}{\Rightarrow} q \qquad A_4 : A_3 \overset{d_1}{\Rightarrow} p$$

$$B_1 : r \qquad B_2 : B_1 \overset{d_2}{\Rightarrow} p \qquad B_3 : B_1 \overset{d_5}{\Rightarrow} \neg q.$$

See Fig. 6 for a graphical representation. Where $\mathrm{AT}_6 = \langle AS_6, \mathcal{K}_6\rangle$ is denoted by $\mathcal{AF}(\mathrm{AT}_6)$. Note that $\mathsf{Prf}(\mathcal{AF}(\mathrm{AT}_6)) = \{\mathcal{E}_1, \mathcal{E}_2\} = \{\{A_1,A_2,A_3,A_4\}, \{A_1,B_1,B_2,B_3\}\}$ and that $\mathsf{AllArgs}(p) = \{A_4, B_2\}$. Then:

- $\mathsf{PrfAcc}^{\cup}(p,\mathcal{E}_1) = \{\langle A_4, \{A_2,A_3\}\rangle\}$ and $\mathsf{PrfAcc}^{\cup}(p,\mathcal{E}_2) = \{\langle B_2, \{B_1\}\rangle\}$.
- $\mathsf{PrfAcc}^{\cup}(q,\mathcal{E}_1) = \{\langle A_3, \{A_2,A_3\}\rangle\}$ and $\mathsf{PrfAcc}^{\cup}(\neg q,\mathcal{E}_2) = \{\langle B_3, \{B_1,B_3\}\rangle\}$.

So p is credulously accepted under Prf, because it is the conclusion of argument A_4 and A_4 is defended by A_2 and A_3 in extension $\mathcal{E}_1 = \{A_1,A_2,A_3,A_4\}$ and p is also the conclusion of argument B_2, which is defended by B_1 in the extension $\mathcal{E}_2 = \{A_1,B_1,B_2,B_3\}$. Moreover, q is credulously accepted under Prf because it is the conclusion of argument A_3 which is defended by the arguments A_2 and A_3 in $\mathcal{E}_1$. Furthermore, $\neg q$ is credulously accepted under Prf because it is the conclusion of argument B_3 which is defended by B_1 and B_3 in $\mathcal{E}_2$.

5.2.2. Explanations for non-accepted formulas

In [16] two types of non-acceptance explanations were considered: a non-derivability explanation and a non-acceptance explanation. The first is aimed at formulas within the language that are not derivable given the argumentation theory. In particular, for a formula ϕ, this explanation collects the rules for ϕ, the antecedents for these rules and the knowledge base elements that are missing to derive ϕ with these rules [16, Definition 13]. We will not repeat that definition, but rather assume that some argument for ϕ exists (i.e., $\mathsf{AllArgs}(\phi) \neq \emptyset$) and focus on the second type of explanations.

When a formula ϕ is not accepted even though it is derivable, the explanation contains the arguments for the formula and the reasons why these are not accepted (i.e., the non-acceptance argument explanations from Section 3.2). Thus, the question why some formula ϕ is not accepted is then answered by an explanation of the form 'because even though argument A has ϕ as its conclusion, A is not accepted because it is not defended against attackers $B_1, \ldots, B_n$'.[11]

Definition 23 *(Non-acceptance formula explanation).* Let $\mathcal{AF}(\mathrm{AT})$ be an argumentation framework based on the argumentation theory AT, let $\phi \in \mathcal{L}$ be such that ϕ is not $\star$-accepted and $\mathsf{AllArgs}(\phi) \neq \emptyset$ and let $\mathcal{E}_\phi \in \mathsf{SemWithout}(\phi)$, given some $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$. Then:

$$\mathsf{SemNotAcc}^{\cap}(\phi, \mathcal{E}_\phi) = \left\{ \left\langle A, \mathbb{D}(A, \mathcal{E}_\phi) \right\rangle \mid A \in \mathsf{AllArgs}(\phi) \text{ and } \mathsf{Conc}(A) = \phi \right\}$$

$$\mathsf{SemNotAcc}^{\cup}(\phi) = \{ \langle A, \mathbb{D}(A, \mathcal{E}) \rangle \mid A \in \mathsf{AllArgs}(\phi) \text{ and } \mathcal{E} \in \mathsf{SemWithout}(\phi) \}.$$

Recall that a formula ϕ is not skeptically accepted if there is some Sem-extension without an argument for ϕ, while ϕ is not credulously accepted if no Sem-extension contains an argument for ϕ. In the above definition, this is formulated by returning for the (skeptical) $\cap$-explanation for the given extension $\mathcal{E}_\phi$ for each argument A with conclusion ϕ the pair consisting of A and $\mathbb{D}(A, \mathcal{E}_\phi)$ (why is ϕ not skeptically accepted under Sem or, in other words, which arguments for ϕ exist and to which attacks on these arguments does the given extension $\mathcal{E}_\phi$ not provide a defense?). The (credulous) $\cup$-explanation collects all the pairs containing an argument A for ϕ and $\mathbb{D}(A, \mathcal{E})$, where $\mathcal{E}$ is some Sem-extension why is ϕ not credulously accepted under Sem or, in other words, which arguments for ϕ exist and to which arguments is no defense provided for every extension?). Since no argument for ϕ is part of any Sem-extension in the case of credulous non-acceptance, all pairs of A and $\mathcal{E}$ are collected in the $\cup$-explanation.

Example 26. Recall from Example 21, that in the argumentation framework $\mathcal{AF}(\mathrm{AT}_1)$ from Example 20, none of the arguments from Example 18 are not credulously accepted, but all arguments shown in Fig. 1 are not skeptically accepted, for $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Prf}, \mathsf{Stb}, \mathsf{Sstb}\}$. Hence, for the same Sem:

- $\mathsf{SemNotAcc}^{\cap}(u, \mathcal{E}_4) = \{\langle A, \{B, D, F\}\rangle\}$ since A is the only argument for u and A is not part of the Sem-extension $\mathcal{E}_4 = \{B, D, F\}$;
- $\mathsf{SemNotAcc}^{\cap}(\neg n(d_1), \mathcal{E}_1) = \{\langle B, \{C, E\}\rangle\}$, $\mathsf{SemNotAcc}^{\cap}(\neg n(d_1), \mathcal{E}_2) = \{\langle B, \{C\}\rangle\}$ and $\mathsf{SemNotAcc}^{\cap}(\neg n(d_1), \mathcal{E}_3) = \{\langle B, \{E\}\rangle\}$ since B is the only argument for $\neg n(d_1)$ and B is not part of the Sem-extensions $\mathcal{E}_1 = \{A, C, E\}$, $\mathcal{E}_2 = \{A, C, F\}$ and $\mathcal{E}_3 = \{A, D, E\}$.

So u is not accepted, even though argument A for u exists, because it is not defended against B, D and F in $\mathcal{E}_4 = \{B, D, F\}$). Similarly, $\neg n(d_1)$ is not skeptically accepted, even though argument B for $\neg n(d_1)$ exists, because it is not defended against C, E or C and E in $\mathcal{E}_2 = \{A, C, F\}$, $\mathcal{E}_3 = \{A, D, E\}$ and $\mathcal{E}_1 = \{A, C, E\}$ respectively).

Example 27. For the argumentation framework $\mathcal{AF}(\mathrm{AT}_6)$ from Example 25, we have that all arguments, except for A_1 are not skeptically accepted. However, if we look at the conclusions, p is skeptically accepted and therefore not non-accepted. We do however, have:

- $\mathsf{PrfNotAcc}^{\cap}(q, \mathcal{E}_2) = \{\langle A_3, \{B_1, B_3\}\rangle\}$;
- $\mathsf{PrfNotAcc}^{\cap}(\neg q, \mathcal{E}_1) = \{\langle B_3, \{A_2, A_3\}\rangle\}$.

So q is not skeptically accepted, even though argument A_3 exists, because it is not defended against the attacks by B_1 and B_3 in $\mathcal{E}_2 = \{A_1, B_1, B_2, B_3\}$) and $\neg q$ is not skeptically accepted, even though argument B_3 exists, because it is not defended against the attacks by A_2 and A_3 in $\mathcal{E}_1 = \{A_1, A_2, A_3, A_4\}$.

5.3. Element explanations

In Section 5.2, explanations are still in terms of arguments. It is further possible to refine these explanations and provide them in terms of elements of arguments – for example, 'formula ϕ is accepted because it is the conclusion of an accepted argument with premises ψ and χ'. To this end we introduce a second function $\mathbb{F}$, which determines the types of elements in the explanation. As instantiations of $\mathbb{F}$ some of the notations from Definition 17 will already be useful. For example, with $\mathbb{F} = \mathsf{Prem}$, the explanations will be in terms of the premises (i.e., the knowledge base elements) of the arguments from the basic explanations and with $\mathbb{F} = \mathsf{Rules}$ the explanations will be in terms of the rules with which the arguments from the basic explanations were constructed. See [16] for additional example instantiations.

The function $\mathbb{F}$ is applied over the elements of the formula explanations from the previous section. For the existence explanation, suppose that ϕ is $\star$-accepted for $\mathsf{Sem} \in \{\mathsf{Nav}, \mathsf{Stg}, \mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$, then:

[11] Note this is for the completeness-based semantics NoDefAgainst interpretation of $\mathbb{D}$, which we use as standard throughout the paper.

$$\mathsf{SemExist}^{\star}(\phi) = \{\mathbb{F}(A) \mid A \in \mathsf{SemAccept}(\phi)\}.$$

So in addition to returning the arguments for ϕ (Definition 21), the existence explanations can now also return the premises or rules used in the construction of those arguments by varying $\mathbb{F}$.

Definition 24 *(Element explanations).* Let $\mathcal{AF}(\mathrm{AT}) = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework based on the argumentation theory AT and let $\phi \in \mathcal{L}$ be such that $\mathsf{AllArgs}(\phi) \neq \emptyset$. For the acceptance explanation, suppose that ϕ is $\star$-accepted for $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$, let $\mathcal{E}_\phi \in \mathsf{Sem}(\mathcal{AF})$ such that $\phi \in \mathsf{Concs}(\mathcal{E}_\phi)$ then:

$$\mathsf{SemAcc}^{\cap}(\phi) = \{\langle \mathbb{F}(A), \mathbb{F}(\mathbb{D}(A, \mathcal{E}))\rangle \mid A \in \mathsf{SemExist}^{\cap}(\phi) \text{ and } \mathcal{E} \in \mathsf{SemWith}(A)\}$$

$$\mathsf{SemAcc}^{\cup}(\phi, \mathcal{E}_\phi) = \left\{\left\langle \mathbb{F}(A), \mathbb{F}(\mathbb{D}(A, \mathcal{E}_\phi))\right\rangle \mid A \in \mathcal{E}_\phi \text{ and } \mathsf{Conc}(A) = \phi\right\}.$$

Now, for the non-acceptance explanation, suppose that ϕ is $\star$-non-accepted for $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$ while $\mathsf{AllArgs}(\phi) \neq \emptyset$ and let $\mathcal{E}_\phi \in \mathsf{SemWithout}(\phi)$ then:

$$\mathsf{SemNotAcc}^{\cap}(\phi, \mathcal{E}_\phi) = \left\{\left\langle \mathbb{F}(A), \mathbb{F}(\mathbb{D}(A, \mathcal{E}_\phi))\right\rangle \mid A \in \mathsf{AllArgs}(\phi) \text{ and } \mathsf{Conc}(A) = \phi\right\}$$

$$\mathsf{SemNotAcc}^{\cup}(\phi) = \{\langle \mathbb{F}(A), \mathbb{F}(\mathbb{D}(A, \mathcal{E}))\rangle \mid A \in \mathsf{AllArgs}(\phi) \text{ and } \mathcal{E} \in \mathsf{SemWithout}(\phi)\}.$$

The above definition applies to each of the explanations from Definitions 22 and 23 the function $\mathbb{F}$.[12] So, for example, in the case of the credulous $\cup$-explanation, $\mathbb{F} = \mathsf{Prem}$ and $\mathbb{D} = \mathsf{Defending}$, it returns the premises of each argument A for ϕ together with the premises of the arguments that defend A in the given extension $\mathcal{E}_\phi$ (which knowledge base elements are required to infer and defend ϕ in $\mathcal{E}_\phi$?). Intuitively, the explanations in the previous section can be understood as explanations as defined here, with $\mathbb{F} = \mathsf{id}$, where $\mathsf{id}(\mathsf{S}) = \mathsf{S}$ for any set S. Note that, while in the above definition $\mathbb{F}$ is the same for both elements, a user might decide to apply two different instantiations of $\mathbb{F}$ within the same explanation.

Remark 10. Algorithm 1 can again be applied to calculate formula explanations, if it can be applied for the chosen $\mathbb{D}$ instantiation. First, note that a formula explanation results in a pair $\langle \mathbb{F}(A), \mathbb{F}(\mathsf{S})\rangle$. Then the algorithm helps to determine S, in the same way as it did in Sections 3.4 and 4. Applying $\mathbb{F}$ to S can then be done by extracting the argument information from the elements of S.

Example 28. For our running example with $\mathcal{AF}(\mathrm{AT}_1)$, recall that $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_1) = \{\langle u, \{C, E\}\rangle\}$, $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_2) = \{\langle A, \{C\}\rangle\}$ and $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_3) = \{\langle A, \{E\}\rangle\}$ as well as $\mathsf{PrfAcc}^{\cup}(\neg n(d_1), \mathcal{E}_4) = \{\langle B, \{D, F\}\rangle\}$, for the non-acceptance of u: $\mathsf{PrfNotAcc}^{\cap}(u, \mathcal{E}_4) = \langle A, \{B, D, F\}\rangle$ and the non-acceptance of $\neg n(d_1)$: $\mathsf{PrfNotAcc}^{\cup}(\neg n(d_1), \mathcal{E}_1) = \{\langle B, \{C, E\}\rangle\}$, $\mathsf{PrfNotAcc}^{\cup}(\neg n(d_1), \mathcal{E}_2) = \{\langle B, \{C\}\rangle\}$ and $\mathsf{PrfNotAcc}^{\cup}(\neg n(d_1), \mathcal{E}_3) = \{\langle B, \{E\}\rangle\}$, for $\mathbb{D}^{\mathsf{acc}} = \mathsf{Defending}$ and $\mathbb{D}^{\mathsf{nacc}} = \mathsf{NoDefAgainst}$. By varying $\mathbb{F}$ we obtain:

1. $\mathsf{PrfExist}^{\cup}(u) = \{w\}$ for $\mathbb{F} = \mathsf{Prem}$;
2. $\mathsf{PrfExist}^{\cup}(\neg n(d_1)) = \{d_2, d_3, d_4\}$ for $\mathbb{F} = \mathsf{Rules}$;
3. $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_1) = \{\langle\{w\}, \{r, s, t\}\rangle\}$, $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_2) = \{\langle\{w\}, \{r, s\}\rangle\}$ and $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_3) = \langle\{w\}, \{r, t\}\rangle\}$ for $\mathbb{F} = \mathsf{Prem}$ and $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_1) = \{\langle\{d_1\}, \{d_3, d_5\}\rangle\}$, $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_2) = \{\langle\{d_1\}, \{d_3\}\rangle\}$ and $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_3) = \{\langle\{d_1\}, \{d_3, d_5\}\rangle\}$ for $\mathbb{F} = \mathsf{Rules}$;
4. $\mathsf{PrfAcc}^{\cup}(\neg n(d_1), \mathcal{E}_4) = \{\langle\{v\}, \{v\}\rangle\}$ for $\mathbb{F} = \mathsf{Prem}$ and $\mathsf{PrfAcc}^{\cup}(\neg n(d_1), \mathcal{E}_4) = \{\langle\{d_2, d_4, d_6\}, \{d_4, d_6\}\rangle\}$ for $\mathbb{F} = \mathsf{Rules}$;
5. $\mathsf{PrfNotAcc}^{\cap}(u, \mathcal{E}_4) = \{\langle\{w\}, \{v\}\rangle\}$ for $\mathbb{F} = \mathsf{Prem}$ and $\mathsf{PrfNotAcc}^{\cap}(u, \mathcal{E}_4) = \{\langle\{d_1\}, \{d_2, d_4, d_6\}\rangle\}$ for $\mathbb{F} = \mathsf{Rules}$;
6. $\mathsf{PrfNotAcc}^{\cup}(\neg n(d_1), \mathcal{E}_1) = \{\langle\{v\}, \{r, s, t\}\rangle\}$, $\mathsf{PrfNotAcc}^{\cup}(\neg n(d_1), \mathcal{E}_2) = \{\langle\{v\}, \{r, s\}\rangle\}$ and $\mathsf{PrfNotAcc}^{\cup}(\neg n(d_1), \mathcal{E}_3) = \langle\{v\}, \{r, t\}\rangle\}$ for $\mathbb{F} = \mathsf{Prem}$ and $\mathsf{PrfNotAcc}^{\cup}(\neg n(d_1), \mathcal{E}_1) = \{\langle\{d_2, d_4, d_6\}, \{d_3, d_5\}\rangle\}$, $\mathsf{PrfNotAcc}^{\cup}(\neg n(d_1), \mathcal{E}_2) = \{\langle\{d_2, d_4, d_6\}, \{d_3\}\rangle\}$ and finally for $\mathcal{E}_3$: $\mathsf{PrfNotAcc}^{\cup}(\neg n(d_1), \mathcal{E}_3) = \{\langle\{d_2, d_4, d_6\}, \{d_3, d_5\}\rangle\}$ for $\mathbb{F} = \mathsf{Rules}$.

The first item, $\mathsf{PrfExist}^{\cup}(u)$ with $\mathbb{F} = \mathsf{Prem}$ can then be put into words as follows: 'u is credulously accepted under Prf because there is an argument for u with premise w.' Explanation $\mathsf{PrfAcc}^{\cup}(u, \mathcal{E}_1)$ from item 3 is one tuple which reads as 'u is credulously accepted under Prf because there is an argument for u with premise w, and this argument is defended against its attackers by one or more arguments with premises r, s, t.' Finally, the first explanation from item 5 can be read as 'u is not skeptically accepted under Prf in $\mathcal{E}_4$ because even though there is an argument for u with premise w, this argument is not defended against one or more attacking arguments with premise v.'

Example 29. For the argumentation framework $\mathcal{AF}(\mathrm{AT}_6)$, introduced in Example 25, some of the element explanations, where $\mathbb{D}^{\mathsf{acc}} = \mathsf{Defending}$ and $\mathbb{D}^{\mathsf{nacc}} = \mathsf{NoDefAgainst}$, are:

- $\mathsf{PrfExist}^{\cup}(p) = \{t, r\}$ for $\mathbb{F} = \mathsf{Prem}$ and $\mathsf{PrfExist}^{\cup}(p) = \{d_1, d_2, d_3, d_4\}$ for $\mathbb{F} = \mathsf{Rules}$;
- $\mathsf{PrfAcc}^{\cup}(p, \mathcal{E}_1) = \{\langle\{t\}, \{t\}\rangle\}$ and $\mathsf{PrfAcc}^{\cup}(p, \mathcal{E}_2) = \{\langle\{r\}, \{r\}\rangle\}$ for $\mathbb{F} = \mathsf{Prem}$ and for $\mathbb{F} = \mathsf{Rules}$: $\mathsf{PrfAcc}^{\cup}(p, \mathcal{E}_1) = \{\langle\{d_1, d_3, d_4\}, \{d_3, d_4\}\rangle\}$ and $\mathsf{PrfAcc}^{\cup}(p, \mathcal{E}_2 = \{\langle\{d_2\}, \emptyset\rangle\}$.

[12] We overload here the notation of the explanations from the previous section. We do so to avoid further notation and since the explanations from the previous section can be seen as a special case of the definition here.

- $\mathsf{PrfNotAcc}^{\cap}(q, \mathcal{E}_2) = \{\langle\{t\}, \{t\}\rangle\}$ for $\mathbb{F} = \mathsf{Prem}$ and $\mathsf{PrfNotAcc}^{\cap}(\neg q, \mathcal{E}_1) = \{\langle\{d_5\}, \{d_3, d_4\}\rangle\}$ for $\mathbb{F} = \mathsf{Rules}$.

An illustration of the possibilities for $\mathbb{D}$ and $\mathbb{F}$ in the context of a real-life application will be discussed in Section 7.

6. Minimality, necessity and sufficiency

We have now provided a flexible and generic framework for explaining the (non-)acceptance of an argument with relevant arguments under different semantics. We have shown how the notion of relevance can be varied by taking different instantiations of $\mathbb{D}$, and how the structure of arguments can be taken into account when selecting both the explanandum and the explanations. We now turn to our second main contribution: taking into account how humans select relevant explanations. Miller [46] mentions many possible selection criteria for explanations. We focus on three, namely minimality, necessity and sufficiency.

- *Minimality* selects explanations that contain fewer elements [62]. In the context of argumentation we interpret this as (subset)minimality in Section 6.1.
- *Necessity* and *sufficiency* provide an explanation in terms of elements (e.g., causes) that are necessary or sufficient for the explanandum [42,43,66]. In the context of argumentation: which arguments are necessary or sufficient for the (non-)acceptance of the argument or formula that is to be explained? Necessity and sufficiency are approached differently for acceptance and non-acceptance and we will hence discuss these separately in Section 6.2 and 6.3, respectively.

This section is an extension of [17]. In particular, we add the notion of minimality and study its relation to necessity and sufficiency in more detail. Moreover, we consider both skeptical and credulous acceptance and discuss how formula explanations can be refined with $\mathbb{F}$.

6.1. Minimality

Minimality is a common way of choosing between explanations in, for example, formal models of causal diagnosis (cf. [22]). As was shown in the previous sections (i.e., Sections 3.3, 4 and 5.3), the size of an explanation can already be reduced with the choice of the semantics and the functions $\mathbb{D}$ and $\mathbb{F}$. For example, an explanation for grounded semantics never contains more arguments than an explanation for any of the other semantics (Propositions 1 and 4 and Corollaries 1 and 3) and explanations with $\mathbb{D} = \mathsf{DirDefending}$ can be smaller than explanations with $\mathbb{D} = \mathsf{Defending}$ (Remark 9). In this section, we further focus on imposing explicit mimimality constraints on specific explanations for (non-)acceptance.

Minimality can mean either minimal set size or subset minimality, and has been discussed for argumentation-based explanations before [33,34,40]. The two notions of minimality were introduced in [33]: set size minimality $\leq$, where $\mathsf{S}_1 \leq \mathsf{S}_2$ denotes $|\mathsf{S}_1| \leq |\mathsf{S}_2|$, and subset minimality $\subseteq$, called compactness by Fan and Toni [33]. These notions of minimality can be integrated in our setting as well. Below we define minimal explanations for the $\cup$-variations from Definitions 6, 8 and 24, the $\cap$-explanations are similar and left to the reader:

Definition 25. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework, possibly based on an argumentation theory AT. Let $A \in \mathsf{Args}$ and, in the case that $\mathcal{AF}$ is based on AT, let $\phi \in \mathcal{L}$. Moreover, let $\preceq \in \{\leq, \subseteq\}$. First suppose that A [resp. ϕ] is credulously accepted for $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$. Then:

- $\mathsf{S} \subseteq \mathsf{Args}$ is a $\preceq$-minimal $\cup$-explanation for the acceptance of A iff $\mathsf{S} \in \min^{\preceq}_{\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})} \mathsf{SemAcc}^{\cup}(A, \mathcal{E})$;
- $\mathsf{S} \subseteq \mathsf{Args}$ is a $\preceq$-minimal $\cup$-explanation for the acceptance of ϕ iff $\mathsf{S} \in \min^{\preceq}_{\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})} \mathsf{SemAcc}^{\cup}(\phi, \mathcal{E})$;

Now suppose that A [resp. ϕ] is not credulously accepted for $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$, then:

- $\mathsf{S} \subseteq \mathsf{SemNotAcc}^{\cup}(A)$ is a $\preceq$-minimal $\cup$-explanation for the non-acceptance of A iff $\mathsf{S} = \{\mathsf{T} \mid \mathsf{T} \in \min^{\preceq} \mathsf{SemNotAcc}^{\cup}(A)\}$;
- $\mathsf{S} \subseteq \mathsf{SemNotAcc}^{\cup}(\phi)$ is a $\preceq$-minimal $\cup$-explanation for the non-acceptance of ϕ iff $\mathsf{S} = \{\mathsf{T} \mid \mathsf{T} \in \min^{\preceq} \mathsf{SemNotAcc}^{\cup}(\phi)\}$.

The explanations in this definition only contain the $\leq$-minimal (if $\preceq = \leq$) or $\subseteq$-minimal (if $\preceq = \subseteq$) explanations from the basic explanations presented earlier in the paper.

Example 30. Recall that for the argumentation framework $\mathcal{AF}_1$ from Example 1 we had that $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_1) = \{C, E\}$, $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_2) = \{C\}$ and $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_3) = \{E\}$ as described in Example 6. Of these possible explanations, the explanations for $\mathcal{E}_2$, i.e., $\{C\}$ and for $\mathcal{E}_3$, i.e., $\{E\}$ are both $\subseteq$-minimal and $\leq$-minimal. The explanation for $\mathcal{E}_1$, i.e., $\{C, E\}$ cannot be considered a minimal explanation for both $\leq$ and $\subseteq$.

In [16, Proposition 2] it was shown that the explanations from [33] correspond to the set of ∪-explanations from Definition 6 for $\mathsf{Sem} = \mathsf{Adm}$ and $\mathbb{D} = \mathsf{Defending}$. We do not repeat the result here, but rather note that, as a result, the minimal explanations as presented here correspond to the minimal explanations in [33].

6.2. Necessity and sufficiency for acceptance

Two important criteria when selecting explanations are necessity and sufficiency (cf., e.g., [42,43,66]). A set of causes C is *sufficient* for the explanandum E if no other causes are required, while C is *necessary* for E if, in order for E to occur, C has to happen as well. In the context of logical implication $\rightarrow$, one could model sufficiency by $C \rightarrow E$ and necessity by $E \rightarrow C$ [41]. In the context of (abstract) argumentation, we say that a set of defending arguments S is *sufficient* for the acceptance of an argument A if no other arguments than those in S need to be accepted for A to be accepted, while S is *necessary* for the acceptance of A if A can only be accepted if S is accepted as well. Sufficiency can be understood as a credulous property. For example, while S might be sufficient for the acceptance of A, there might be other S's which would ensure the acceptance of A as well. On the other hand, necessity can be understood as a skeptical property. For example, when S is necessary for the acceptance of A, it is impossible to accept A without accepting all the arguments in S.

Like for the basic framework, we will discuss argument and formula explanations separately. The reason for this will become clear in the sections on formula explanations.

6.2.1. Necessity and sufficiency for accepted arguments

In abstract argumentation, we say that a (conflict-free) set of accepted arguments S is *sufficient* for the acceptance of some argument A if S guarantees, independent of the status of other arguments, that A is accepted by defending A against all its attackers. An (accepted) argument B is *necessary* for the acceptance of argument A if B defends A and it is impossible to accept A without accepting B. In the following definition, as well as later on in this section, we will assume that the arguments in an explanation for an argument A are *relevant* for A. Recall from Definition 4 that this entails that all arguments in an explanation for A (in)directly attack or defend A and that no argument attacks itself.

Definition 26. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and let $A \in \mathsf{Args}$ be accepted (w.r.t. Sem and $\cup$ or $\cap$). Then:

- $\mathsf{S} \subseteq \mathsf{Args}$ is *sufficient for the acceptance* of A if S is relevant for A, S is conflict-free and S defends $\mathsf{S} \cup \{A\}$ against all its attackers;
- $B \in \mathsf{Args}$ is *necessary for the acceptance* of A if B is relevant for A and if $B \notin \mathcal{E}$ for some $\mathcal{E} \in \mathsf{Adm}(\mathcal{AF})$, then $A \notin \mathcal{E}$.

Remark 11. The above definition is aimed at admissibility-based semantics, hence that for sufficiency S has to defend A against all its attackers. For other types of semantics (e.g., based on conflict-freeness), the definition can be adjusted (e.g., by removing the third condition on S from the definition of sufficiency).

Example 31. In the argumentation framework $\mathcal{AF}_1$ both $\{C\}$ and $\{E\}$ are sufficient for the acceptance of A but neither is necessary, while for the acceptance of B, $\{D, F\}$ is sufficient and both D and F are necessary.

The notions of minimality from Section 6.1 can also be applied to sufficient sets (note that an argument is either necessary for the acceptance of some argument or not, therefore, minimality is not relevant for necessity).

Definition 27. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and let $A \in \mathsf{Args}$ be accepted (w.r.t. Sem and $\cup$ or $\cap$). Then:

- $\mathsf{S} \subseteq \mathsf{Args}$ is *$\leq$-minimally sufficient for the acceptance of* A if S is sufficient for the acceptance of A and there is no $\mathsf{S}' \subseteq \mathsf{Args}$ that is sufficient for the acceptance of A such that $|\mathsf{S}'| \leq |\mathsf{S}|$;
- $\mathsf{S} \subseteq \mathsf{Args}$ is *$\subseteq$-minimally sufficient for the acceptance of* A if S is sufficient for the acceptance of A and there is no $\mathsf{S}' \subseteq \mathsf{Args}$ that is sufficient for the acceptance of A such that $\mathsf{S}' \subset \mathsf{S}$.

We can now define the above notions of (minimal) sufficiency and necessity as variations of $\mathbb{D}$.

Definition 28. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and let $A \in \mathsf{Args}$ be accepted (w.r.t. Sem and $\cup$ or $\cap$). Then[13]:

- $\mathsf{Suff}(A, \emptyset) = \{\mathsf{S} \cup \{A\} \subseteq \mathsf{Args} \mid \mathsf{S} \text{ is sufficient for the acceptance of } A\}$ denotes the set of all sufficient sets of arguments for the acceptance of A;
- $\mathsf{MinSuff}^{\preceq}(A, \emptyset) = \mathsf{mima}\ \mathsf{Suff}(A, \emptyset)$, where $\preceq = \leq$ if $\mathsf{mima} = \min^{\leq}$ and $\preceq = \subseteq$ if $\mathsf{mima} = \min^{\subseteq}$, denotes the set of all $\preceq$-minimally sufficient sets of arguments for the acceptance of A;

[13] Suff and Nec are defined for an argument and some set of arguments, not referred to in the definition, therefore the second argument of the function is empty. We do so because $\mathbb{D}$ requires an argument and a set of arguments.

- $\mathsf{Nec}(A,\emptyset) = \{B \in \mathsf{Args} \mid B \text{ is necessary for the acceptance of } A\} \cup \{A\}$ denotes the set of all arguments that are necessary for the acceptance of A.

Remark 12. When $\mathbb{D} \in \{\mathsf{Suff}, \mathsf{MinSuff}^{\preceq}, \mathsf{Nec}\}$, $\mathsf{SemAcc}^{\star}$ is the same for any semantics Sem. This is the case since the definition of sufficiency and necessity is not defined w.r.t. Sem. Recall from Definition 5 that $\mathsf{Defending}(A,\mathcal{E}) = \mathsf{Defending}(A) \cap \mathcal{E}$, this is possible since $\mathsf{Defending}(A)$ contains all the arguments that defend A. If we would take some sufficient set and take the intersection with an arbitrary extension, this might lead to meaningless explanations. For example, as argued in Example 31, $\{E\}$ is sufficient for the acceptance of A, but if we would compare it with the preferred extension $\{A,C,F\}$, the explanation would not be relevant, as E is not relevant for the acceptance of A in this extension.

Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an AF, $A \in \mathsf{Args}$ be a credulously accepted argument and let $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ be any extension. When implementing the above variations of $\mathbb{D}$ in the credulous acceptance explanation from Definition 6 we obtain the following[14]:

- If $\mathbb{D} = \mathsf{Suff}$, then: $\mathsf{SemAcc}^{\cup}(A,\mathcal{E}) = \mathsf{Suff}(A,\emptyset)$;
- If $\mathbb{D} = \mathsf{Nec}$, then $\mathsf{SemAcc}^{\cup}(A,\mathcal{E}) = \mathsf{Nec}(A,\emptyset)$.

The next two examples shows how necessary and sufficient explanations differ from basic explanations as discussed in Sections 3 and 4.

Example 32. For the running example with $\mathcal{AF}_1$ we have that both A and B are credulously accepted when $\mathsf{Sem} = \mathsf{Prf}$.

- If $\mathbb{D} = \mathsf{Suff}$, then $\mathsf{PrfAcc}^{\cup}(A,\mathsf{S}) = \{\{A,C\},\{A,E\},\{A,C,E\}\}$, while if $\mathbb{D} = \mathsf{MinSuff}^{\preceq}$, then $\mathsf{PrfAcc}^{\cup}(A,\mathsf{S}) = \{\{A,C\},\{A,E\}\}$, where $\mathsf{S} \in \{\mathcal{E}_1,\mathcal{E}_2,\mathcal{E}_3\}$. That the extension does not influence the explanation follows since the functions from Definition 28 do not rely on a specific extension. These explanations show that there are three sufficient sets for the acceptance of A, containing at least A and C or E and that two of these sets are also minimally sufficient.
- If $\mathbb{D} = \mathsf{Nec}$, then $\mathsf{PrfAcc}^{\cup}(A,\mathsf{S}) = \{A\}$ for $\mathsf{S} \in \{\mathcal{E}_1,\mathcal{E}_2,\mathcal{E}_3\}$. For the acceptance of A only A is necessary. This is the case because A is added to the set of necessary arguments.
- If $\mathbb{D} \in \{\mathsf{Suff}, \mathsf{MinSuff}^{\preceq}\}$, then $\mathsf{PrfAcc}^{\cup}(B,\mathcal{E}_4) = \{\{B,D,F\}\}$. Thus, there is only one sufficient set for the acceptance of B containing B, D and F, this set is also minimal.
- If $\mathbb{D} = \mathsf{Nec}$, then $\mathsf{PrfAcc}^{\cup}(B,\mathcal{E}_4) = \{B,D,F\}$. Thus, for the acceptance of B all three arguments B, D and F are necessary.

Note that these explanations correspond to the intuitions in Example 31.

Example 33. Recall the argumentation framework $\mathcal{AF}_2$ from Example 7, shown in Fig. 2. Recall that $\mathsf{Prf}(\mathcal{AF}_2) = \{\mathcal{E}_1,\mathcal{E}_2\} = \{\{A,C,D\},\{A,C,E\}\}$.

- If $\mathbb{D} = \mathsf{Suff}$, then $\mathsf{PrfAcc}^{\cup}(A,\mathsf{S}) = \{\{A,C\},\{A,D\},\{A,C,D\}\}$ for $\mathsf{S} \in \{\mathcal{E}_1,\mathcal{E}_2\}$.
- If $\mathbb{D} = \mathsf{Nec}$, then $\mathsf{PrfAcc}^{\cup}(A,\mathsf{S}) = \emptyset$ for $\mathsf{S} \in \{\mathcal{E}_1,\mathcal{E}_2\}$.

Thus, even though C is always part of the extension under any completeness-based semantics, $\{A,D\}$ is sufficient for the acceptance of A. Therefore, C is not necessary for the acceptance of A either, since the acceptance of D is already sufficient.

6.2.2. Necessity and sufficiency for accepted formulas

In Section 5 we started by introducing explanations for formulas (Section 5.2) and then added the possibility to vary the content of an explanation by introducing the function $\mathbb{F}$ (Section 5.3). By taking $\mathbb{F} = \mathsf{id}$, the explanations in both sections are equivalent. Therefore, in this section we skip the first step and immediately add the function $\mathbb{F}$, where possibly $\mathbb{F} = \mathsf{id}$.

For accepted formulas we start with existence explanations. In particular, we define necessary and sufficient existence explanations for a given formula:

Definition 29. Let $\mathcal{AF}(\mathrm{AT}) = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework based on the argumentation theory AT and let $\phi \in \mathcal{L}$ be such that ϕ is $\star$-accepted for $\mathsf{Sem} \in \{\mathsf{Nav}, \mathsf{Stb}, \mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$:

$$\mathsf{SuffSemExist}^{\star}(\phi) = \{\mathbb{F}(A) \mid A \in \mathsf{SemAccept}(\phi)\}$$

$$\mathsf{NecSemExist}^{\star}(\phi) = \bigcap \{\mathbb{F}(A) \mid A \in \mathsf{SemAccept}(\phi)\}.$$

These existence explanations are based on the ones introduced in Section 5. For the sufficient explanations all the elements based on $\mathbb{F}$ are collected, from all accepted arguments. For the necessary explanations the intersection of these elements is taken: if an

[14] The $\cap$-explanations are similar and left to the reader.

element is part of all the arguments in the existence explanations, it is necessary for the existence of an argument for ϕ. Here it is already interesting to consider other instantiations of $\mathbb{F}$ than id. When $\mathbb{F}$ is instantiated with, e.g., Prem or Rules, the common elements of all the arguments for ϕ can be collected. By varying $\mathbb{F}$ the required elements for the existence of an argument for ϕ can be given as the explanation. Like in the existence explanation introduced in Section 5, the assumption that ϕ is $\star$-accepted is essential.

Example 34. Take an argumentation system with $\mathcal{K}_p = \{r, s, t\}$ and $\mathcal{R} = \{d_1, d_2\}$ such that in addition to the three arguments for r, s and t the following arguments can be derived: $A : r, s \overset{d_1}{\Rightarrow} u$ and $B : s, t \overset{d_2}{\Rightarrow} u$. Here, all arguments are skeptically/credulously accepted, so:

- $\mathsf{SuffPrfExist}^{\star}(u) = \{A, B\}$ and $\mathsf{NecPrfExist}^{\star}(u) = \emptyset$ for $\mathbb{F} = \mathsf{id}$ since A and B are both arguments for u;
- $\mathsf{SuffPrfExist}^{\star}(u) = \{\{r, s\}, \{s, t\}\}$ for $\mathbb{F} = \mathsf{Prem}$ since argument A is constructed with the premises r and s and argument B is constructed with the premises s and t;
- $\mathsf{NecPrfExist}^{\star}(u) = \{s\}$ for $\mathbb{F} = \mathsf{Prem}$ since both arguments for u require the premise s.

We now turn to acceptance explanations. To start with, one can replace $\mathbb{D}$ in Definition 24 with Nec or Suff (replacement with MinSuff is similar to replacement with Suff and left to the reader). Note that $\mathsf{Nec}(A, \mathsf{S})$ is unique for a given argumentation framework $\mathcal{AF}$, there is therefore no difference between the $\cap$-explanation and the $\cup$-explanation. We then obtain, where $A \in \mathsf{SemAccept}(\phi)$ and $\mathsf{S}_\phi \in \mathsf{Suff}(A, \emptyset)$, for example:

$$\mathsf{SemAcc}(\phi) = \{\langle \mathbb{F}(A), \mathbb{F}(\mathsf{S})\rangle \mid A \in \mathsf{SemExist}^{\cap}(\phi) \text{ and } \mathsf{S} \in \mathsf{Nec}(A, \emptyset)\};$$

$$\mathsf{SemAcc}^{\cap}(\phi) = \{\langle \mathbb{F}(A), \mathbb{F}(\mathsf{S})\rangle \mid A \in \mathsf{SemExist}^{\cap}(\phi) \text{ and } \mathsf{S} \in \mathsf{Suff}(A, \emptyset)\};$$

$$\mathsf{SemAcc}^{\cup}(\phi, \mathsf{S}_\phi) = \left\{\left\langle \mathbb{F}(A_\phi), \mathbb{F}(\mathsf{S}_\phi)\right\rangle \mid A_\phi \in \mathsf{S}_\phi \text{ and } \mathsf{Conc}(A_\phi) = \phi\right\}.$$

These explanations thus provide pairs containing (elements of) an argument A for ϕ and the (elements of) the arguments that are necessary or sufficient for accepting A. This could be further tailored by, for example, replacing SemExist with the necessary/sufficient existence explanations from Definition 29, thus providing for each pair a necessary/sufficient argument A for ϕ.

As we will show in the next example, instantiating $\mathbb{D}$ with Nec or Suff might not result in the desirable explanations. This is the case because first necessary/sufficient arguments are determined and only then is the structure of these arguments taken into account by varying $\mathbb{F}$. However, as we already saw in Example 34, taking the structure of arguments into account before determining what is necessary/sufficient can result in other explanations that highlight the necessary or sufficient elements for the acceptance of a formula.

Example 35. For the running example with $\mathcal{AF}(\mathrm{AT}_1)$ we have that A is the only argument for u and we had that C, E and C and E were sufficient for the acceptance of A but no argument was necessary for the acceptance of A. Then:

- where $\mathbb{D} = \mathsf{Suff}$: $\mathsf{PrfAcc}^{\cup}(u, \{A, C\}) = \{\langle A, \{A, C\}\rangle\}$, $\mathsf{PrfAcc}^{\cup}(u, \{A, E\}) = \{\langle A, \{A, E\}\rangle\}$ and $\mathsf{PrfAcc}^{\cup}(u, \{A, C, E\}) = \{\langle A, \{A, C, E\}\rangle\}$ for $\mathbb{F} = \mathsf{id}$ and $\mathsf{PrfAcc}^{\cup}(u, \mathsf{S}) = \{\langle \{w\}, \{r, s, t\}\rangle\}$, for $\mathsf{S} \in \{\{A, C\}, \{A, E\}, \{A, C, E\}\}$ for $\mathbb{F} = \mathsf{Prem}$; and
- where $\mathbb{D} = \mathsf{Nec}$: $\mathsf{PrfAcc}(u) = \{\langle A, \{A\}\rangle\}$ for $\mathbb{F} = \mathsf{id}$ and $\mathsf{PrfAcc}(u) = \{\langle \{w\}, \{w\}\rangle\}$ for $\mathbb{F} = \mathsf{Prem}$.

In this example, the necessary acceptance explanation for the formula u is only based on the argument for u itself: the choice for $\mathbb{F}$ does not change this. Intuitively, this is the case because A can be defended by C or E and both are sufficient for this defense. One could argue, however, that the premise r is necessary for the acceptance of u, since it is part of all the defending arguments of u.

Based on the observations in the above example, necessary and (minimally) sufficient explanations for formulas should take into account the structure of the arguments before determining what elements are necessary/sufficient. For sufficiency (i.e., $\mathbb{D} = \mathsf{Suff}$) the explanation will be very similar to the one where $\mathbb{D}$ was instantiated with Suff. But for MinSuff and Nec, the function $\mathbb{F}$ is applied (i.e., the structure of the arguments is taken into account) before minimal sufficiency resp. necessity is determined. This way minimality is based on the elements of the arguments and not the number of arguments themselves and necessity concerns the required elements, not the required arguments.

Definition 30. Let $\mathcal{AF}(\mathrm{AT}) = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework based on the argumentation theory AT and let $\phi \in \mathcal{L}$ be such that ϕ is $\star$-accepted, for $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$. Moreover, let $A_\phi \in \mathsf{SemAccept}(\phi)$ and $\mathsf{S} \in \mathsf{Suff}(A_\phi, \emptyset)$. First the basic sufficient explanations, based on Definition 24:

$$\mathsf{SuffAcc}^{\cap}(\phi) = \{\langle \mathbb{F}(A), \mathbb{F}(\mathsf{S})\rangle \mid A \in \mathsf{SemExist}^{\cap}(\phi) \text{ and } \mathsf{S} \in \mathsf{Suff}(A, \emptyset)\};$$

$$\mathsf{SuffAcc}^{\cup}(\phi, \mathsf{S}) = \{\langle \mathbb{F}(A), \mathbb{F}(\mathsf{S})\rangle \mid A \in \mathsf{S} \text{ and } \mathsf{Conc}(A) = \phi\}.$$

Now suppose that the explanations have to be minimal as well. Then, based on the above and the credulous explanation from Definition 25, where $\preceq \in \{\leq, \subseteq\}$:

$$\mathsf{SuffAcc}^{\cap}_{\preceq}(\phi) = \left\{ \left\langle \mathbb{F}(A), \mathsf{S}' \right\rangle \;\middle|\; A \in \mathsf{SemExist}^{\cap}(\phi) \text{ and } \mathsf{S}' \in \min^{\preceq}_{\mathsf{S} \in \mathsf{Suff}(A,\emptyset)} \mathbb{F}(\mathsf{S}) \right\}$$

$$\mathsf{SuffAcc}^{\cup}_{\preceq}(\phi,\emptyset) = \left\{ \left\langle \mathbb{F}(A), \mathsf{S}' \right\rangle \;\middle|\; \mathsf{S}' \in \min^{\preceq}_{A \in \mathsf{SemExist}^{\cup}(\phi) \& \mathsf{S} \in \mathsf{Suff}(A,\emptyset)} \mathbb{F}(\mathsf{S}) \right\}.$$

Finally, the explanations for $\mathbb{D} = \mathsf{Nec}$ have to be defined:

$$\mathsf{NecAcc}^{\cap}(\phi) = \left\{ \left\langle \mathbb{F}(A), \bigcap_{\mathsf{S} \in \mathsf{Suff}(A,\emptyset)} \mathbb{F}(\mathsf{S}) \right\rangle \;\middle|\; A \in \mathsf{SemExist}^{\cap}(\phi) \right\}$$

$$\mathsf{NecAcc}^{\cup}(\phi,\emptyset) = \left\{ \left\langle \mathbb{F}(A), \bigcap_{\mathsf{S} \in \mathsf{Suff}(A,\emptyset)} \mathbb{F}(\mathsf{S}) \right\rangle \;\middle|\; A \in \mathsf{SemExist}^{\cup}(\phi) \right\}.$$

Note first that the basic sufficient explanations no longer rely on the extension, as was the case in Definitions 22 and 24, since the extension is not relevant for determining sufficiency. The skeptical sufficient explanation consists therefore of a set with for each argument A for ϕ a pair with the elements of A and a sufficient set for the acceptance of A (which elements or arguments are sufficient for the acceptance of an argument for ϕ?). Thus, it gives for ϕ the sufficient reasons for its acceptance, by showing for each accepted argument A for ϕ a sufficient set for the acceptance of A in terms of (the elements of) arguments. The credulous sufficient explanation returns for each argument A with conclusion ϕ in the given sufficient set S the pair consisting of the elements of A and the elements of S, determined by $\mathbb{F}$ (which (elements of) arguments are sufficient for the acceptance of ϕ in the given sufficient set?). Thus, the explanation shows why S is sufficient for the acceptance of ϕ by giving for each argument A for ϕ in S a pair, consisting of A as a reason for the existence of conclusion ϕ and S in terms of (the elements of) the arguments.

The second set of explanations is similar to the basic sufficient explanations, but takes into account minimality in the second element and rather than a set of sufficient arguments the empty set is provided, since minimality has to be determined. In the skeptical explanation, the second element contains the elements of a sufficient set for the acceptance of an argument for ϕ, such that it is minimal among the set of elements of such sufficient sets (for all arguments of ϕ what are minimal sufficient reasons for accepting this argument?). Thus, rather than minimality over the sets of sufficient arguments, $\mathbb{F}$ is applied before minimality. While for skeptical explanations all arguments for ϕ are considered and minimality is applied to the sufficient sets for these arguments, for the credulous explanations minimality is applied to all accepted arguments for ϕ and the sufficient sets for these arguments (what is a minimal sufficient set for the acceptance of ϕ?). Thus, while both explanations still provide explanations in terms of pairs, containing an existence and acceptance explanation, in the skeptical explanation the acceptance explanation is minimal with respect to the sufficient sets for the argument for ϕ, while in the credulous explanation the acceptance explanation is minimal with respect to the sufficient sets for all the arguments for ϕ.

For necessity for accepted formulas we use that $\mathsf{Nec}(A,\emptyset) = \bigcap \mathsf{Suff}(A,\emptyset)$ for any accepted argument A, which will be shown in Proposition 7. Note that, since necessity is a skeptical property, there is not much difference between the skeptical and credulous explanation, except the requirement that ϕ is skeptically respectively credulously accepted. We cannot simply define a necessity formula explanation by taking $\mathbb{D} = \bigcap \mathsf{Suff}$. This follows since there might not be any necessary arguments, there might be necessary premises or rules, as illustrated in Example 35. Therefore we applied:

- $\mathsf{Nec}(A,\emptyset) = \bigcap_{\mathsf{S} \in \mathsf{Suff}(A,\emptyset)} \mathbb{F}(\mathsf{S})$, which denotes the set of all elements (determined by $\mathbb{F}$) of the intersection of the sufficient sets for A.

Like before, the explanations are sets of pairs, the first element is an argument A for ϕ and the second contains all the necessary arguments or elements there of for the acceptance of A (what elements are necessary for the acceptance of the accepted arguments for ϕ?).

Example 36. For the framework $\mathcal{AF}(\mathrm{AT}_1)$ recall from Example 35 the explanations for $\mathbb{F} = \mathsf{id}$. For $\mathbb{F} = \mathsf{Prem}$ we have now the following explanations:

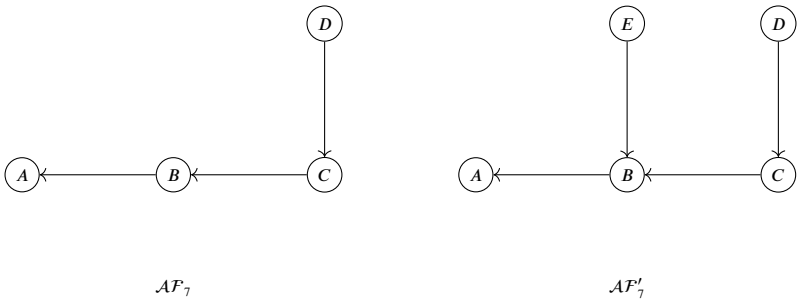
- $\mathsf{SuffAcc}^{\cup}(u) \in \{\langle \{w\}, \{r,s,w\}\rangle, \langle \{w\}, \{r,t,w\}\rangle\}$;
- $\mathsf{NecAcc}^{\cup}(u) = \langle \{w\}, \{r,w\}\rangle$.

So u is credulously accepted, because an argument for u can be constructed from the premise w and there are arguments that sufficiently defend it which are constructed from the premises r and s and from r and t. The premise r is therefore necessary in the defense of the argument for u.

Since the structure of the arguments is taken into account before the intersection of the sufficient explanations is taken, the necessary acceptance explanation for u no longer only relies on the argument A.

6.2.3. *Properties of necessity and sufficiency for acceptance*

Next we show some useful properties of sufficient and necessary (sets of) arguments for acceptance. We show that sufficient sets are admissible, which means that they are conflict-free and defend their own arguments, we discuss conditions under which sufficient



$\mathcal{AF}_7$ $\mathcal{AF}'_7$

Fig. 7. Graphical representation of the argumentation frameworks $\mathcal{AF}_7$ and $\mathcal{AF}'_7$.

and necessary sets are empty and how necessary sets are related to sufficient sets. These results provide insight into the content of sufficient and necessary sets and how these are related. Moreover, based on these results, it can be determined what additional explanations need to be implemented to cover empty explanations. The proofs of the results in this section can be found in A.4.

Proposition 7. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an argumentation framework and let $A \in Args$ be accepted w.r.t. some* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ *and $\star \in \{\cap, \cup\}$. Then:*

1. *For all* $\mathsf{S} \in \mathsf{Suff}(A, \emptyset)$, $\mathsf{S} \in \mathsf{Adm}(\mathcal{AF})$*;*
2. $\mathsf{Suff}(A, \emptyset) = \{A\}$ *iff there is no* $B \in Args$ *such that* $(B, A) \in Att$ *or* $\{A\}$ *defends itself against all its attackers;*
3. $\mathsf{Nec}(A, \emptyset) = \{A\}$ *iff there is no* $B \in Args$ *such that* $(B, A) \in Att$*, or* $\{A\}$ *defends itself against all its attackers or it holds that* $\bigcap \mathsf{Suff}(A, \emptyset) = \{A\}$.
4. $\mathsf{Nec}(A, \emptyset) = \bigcap \mathsf{Suff}(A, \emptyset)$.

This proposition shows that the sets in $\mathsf{Suff}(A, \emptyset)$ are admissible and contain all the needed arguments to defend A in the sense of Definition 1 (item 1). Like for the basic explanations (recall Proposition 3), sufficient sets are only empty when the considered argument is not attacked at all (item 2) and necessary sets are only empty when they are not attacked at all, or there are sufficient sets which have an empty intersection (item 3). Finally, necessary sets are the intersection of all the sufficient sets of the argument (item 4).

The next proposition relates the introduced notions of necessity and sufficiency with $\mathbb{D} = \mathsf{Defending}$. In particular, for all extensions $\mathcal{E}$, the first item shows that $\mathsf{Defending}(A, \mathcal{E})$ is a sufficient set of arguments for the acceptance of A and the second item shows that $\mathsf{Defending}(A, \mathcal{E})$ contains the necessary arguments for A. This shows that, although $\mathsf{Defending}$ is only one of many options for $\mathbb{D}$, by defining the basic explanations in Section 3.1 with $\mathsf{Defending}$, these explanations are always sufficient and contain the necessary arguments.

Proposition 8. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an argumentation framework and let $A \in Args$ be accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ *and $\star \in \{\cap, \cup\}$. Then:*

- *for all* $\mathcal{E} \in \mathsf{SemWith}(A)$, $\mathsf{Defending}(A, \mathcal{E}) \in \mathsf{Suff}(A, \emptyset)$*;*
- $\bigcap_{\mathcal{E} \in \mathsf{SemWith}(A)} \mathsf{Defending}(A, \mathcal{E}) = \mathsf{Nec}(A, \emptyset)$.

Remark 13. One might suspect that, based on the above result, necessary and sufficient explanations can be calculated with Algorithm 1. And indeed, the second bullet above shows that this is the case for necessary explanations (recall from Definition 10 that $\mathsf{Defending}(A, \mathcal{E})$ can be calculated with Algorithm 1). However, for sufficient explanations this is not the case. Note that, in our running example, $\{C, F\} \in \mathsf{Suff}(A, \emptyset)$, however, F does not defend A in any extension.

6.3. Necessity and sufficiency for non-acceptance

When looking at the non-acceptance of an argument A, the acceptance of any of its direct attackers is a sufficient explanation. However, other arguments (e.g., some of the indirect attackers) might be sufficient as well. An argument is necessary for the non-acceptance of A, when it is relevant and A is accepted in the argumentation framework without it. In what follows we will assume that $(A, A) \notin \mathsf{Att}$, since otherwise A itself is the reason for its non-acceptance.

Example 37. Let $\mathcal{AF}_7 = \langle \mathsf{Args}_7, \mathsf{Att}_7\rangle$, shown in Fig. 7. Since D is not attacked at all, D is a sufficient explanation for the non-acceptance of A, since by accepting D, A cannot be accepted under admissibility-based semantics. Now consider $\mathcal{AF}'_7 = \langle \mathsf{Args}_7 \cup \{E\}, \mathsf{Att}_7 \cup \{(E, B)\}\rangle$, also shown in Fig. 7. Here A is accepted, even for $\mathsf{Sem} = \mathsf{Grd}$, since E attacks B. The attack from D is therefore no longer sufficient for the non-acceptance of A. In fact, A is no longer non-accepted.

In order to define sufficiency for non-acceptance we need the following definition, which formalizes the above:

Definition 31. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A, B \in \mathsf{Args}$ such that A indirectly attacks B, via $C_1, \ldots, C_n \in \mathsf{Args}$, i.e., $(A, C_1), (C_1, C_2), \ldots, (C_n, B) \in \mathsf{Att}$. It is said that the attack from A on B is *uncontested* if there is no $D \in \mathsf{Args}$ such that $(D, C_{2i}) \in \mathsf{Att}$ for $i \in \{1, \ldots, \frac{n}{2}\}$. It is *contested* otherwise, in which case it is said that the attack from A is contested in C_{2i} and that C_{2i} is the contested argument.

This definition is needed since the acceptance of an indirect attacker might already be sufficient for the non-acceptance of an argument, but not every indirect attacker is sufficient for non-acceptance. In Example 37 this was shown by the attack from E on B, which contested the attack from D on A in B. In our running example we have:

Example 38. For $\mathcal{AF}_1$ from Example 1 we have that the direct attacks from C and E on B are uncontested. Therefore, both E and C can be seen as sufficient for the non-acceptance of B. However, the attacks from D and F on A are contested in B. For D this follows since it defends B, but B is attacked by E and, similarly, F defends B, but B is attacked by C. Hence, although D and F indirectly attack A, by just accepting one, A is not necessarily non-accepted, therefore neither would be sufficient on its own to make A non-accepted.

6.3.1. Necessity and sufficiency for non-accepted arguments

For the definition of necessary for non-acceptance we define subframeworks, which are needed since an argument might be non-accepted since it is attacked by an accepted or by another non-accepted argument.[15]

Definition 32. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A \in \mathsf{Args}$. Then $\mathcal{AF}_{\downarrow A} = \langle \mathsf{Args} \setminus \{A\}, \mathsf{Att} \cap (\mathsf{Args} \setminus \{A\} \times \mathsf{Args} \setminus \{A\}) \rangle$ denotes the AF based on $\mathcal{AF}$ but without A.

Since indirect attacks might be sufficient for not accepting an argument, but they also might be contested, the definition of sufficiency for non-acceptance is defined inductively.

Definition 33. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A \in \mathsf{Args}$ be non-accepted (w.r.t. Sem and $\cup$ or $\cap$). Then:

- $\mathsf{S} \subseteq \mathsf{Args}$ is *sufficient for the non-acceptance* of A if S is relevant for A and there is a $B \in \mathsf{S}$ such that:
  - $(B, A) \in \mathsf{Att}$; or
  - B indirectly attacks A and that attack is uncontested; or
  - B indirectly attacks A and for every argument C in which the attack from B on A is contested and every $D \in \mathsf{Args}$ such that $(D, C) \in \mathsf{Att}$, there is an $\mathsf{S}' \subseteq \mathsf{S}$ that is sufficient for the non-acceptance of D.
- $B \in \mathsf{Args}$ is *necessary for the non-acceptance* of A if B is relevant for A and A is accepted w.r.t. Sem and $\cup$ or $\cap$ in $\mathcal{AF}_{\downarrow B}$.

Example 39. For $\mathcal{AF}_1$ from Example 1 we have that B is both necessary and sufficient for the non-acceptance of A. Moreover, while D and F are neither sufficient for the non-acceptance of A, $\{D, F\}$ is. For the non-acceptance of B we have that C and E are sufficient, but neither of these is necessary.

We can now introduce the above notions as variations of $\mathbb{D}$.

Definition 34. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A \in \mathsf{Args}$ be non-accepted (w.r.t. Sem and $\cup$ or $\cap$). Then:

- $\mathsf{SuffNot}(A, \emptyset) = \{\mathsf{S} \subseteq \mathsf{Args} \mid \mathsf{S} \text{ is sufficient for the non-acceptance of } A\}$, denotes the set of sets of arguments that, when accepted, cause A to be non-accepted;
- $\mathsf{MinSuffNot}^{\preceq}(A, \emptyset) = \{\mathsf{S} \in \mathsf{SuffNot}(A, \emptyset) \mid \nexists \mathsf{S}' \in \mathsf{SuffNot}(A, \emptyset) \text{ such that } \mathsf{S}' \preceq \mathsf{S}\}$, denotes the set of all $\preceq$-minimally sufficient sets for the non-acceptance of A.
- $\mathsf{NecNot}(A, \emptyset) = \{B \in \mathsf{Args} \mid B \text{ is necessary for the non-acceptance of } A\}$, denotes the set of all arguments that are necessary for A not to be accepted.

Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF, $A \in \mathsf{Args}$ be a not credulously accepted argument and let $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ be any extension. When implementing the above variations of $\mathbb{D}$ in the credulous non-acceptance explanation from Definition 8, we obtain the following[16]:

- If $\mathbb{D} = \mathsf{SuffNot}$, then: $\mathsf{SemNotAcc}^{\cup}(A, \mathcal{E}) = \{\mathsf{SuffNot}(A, \emptyset) \mid \mathcal{E} \in \mathsf{SemWithout}(A)\} = \mathsf{SuffNot}(A, \emptyset)$;
- If $\mathbb{D} = \mathsf{NecNot}$, then: $\mathsf{SemNotAcc}^{\cup}(A, \mathcal{E}) = \{\mathsf{NecNot}(A, \emptyset) \mid \mathcal{E} \in \mathsf{SemWithout}(A)\} = \mathsf{NecNot}(A, \emptyset)$.

[15] In terms of labeling semantics (see e.g., [5]) an argument is non-accepted if it is `out` (i.e., attacked by an `in` argument) or `undecided`.

[16] The $\cap$-explanations are similar and left to the reader.

Example 40. For $\mathcal{AF}_1$, let $\mathsf{S} \in \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3\}$ be an extension. Recall that both A and B are not skeptically accepted when $\mathsf{Sem} = \mathsf{Prf}$.

- If $\mathbb{D} = \mathsf{SuffNot}$ then: $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}_4) = \{\{B\}, \{D, F\}, \{B, D, F\}\}$. This means that for the non-acceptance of A there are three sets of sufficient arguments. If $\mathbb{D} = \mathsf{MinSuffNot}^{\subseteq}$ then $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}_4) = \{\{B\}, \{D, F\}\}$ while if $\mathbb{D} = \mathsf{MinSuffNot}^{\leq}$ then $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}_4) = \{B\}$.
- If $\mathbb{D} = \mathsf{NecNot}$ then: $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}_4) = \{B, D, F\}$. There are three necessary arguments for the non-acceptance of A: B, D and F.
- If $\mathbb{D} = \mathsf{SuffNot}$ then: $\mathsf{PrfNotAcc}^{\cap}(B, \mathsf{S}) = \{\{C\}, \{E\}, \{C, E\}\}$ and if $\mathbb{D} = \mathsf{MinSuffNot}^{\preceq}$ then $\mathsf{PrfNotAcc}^{\cap}(B, \mathsf{S}) = \{\{C\}, \{E\}\}$. Thus, there are three sufficient sets for the non-acceptance B, two of which are minimal: $\{C\}$ and $\{E\}$.
- If $\mathbb{D} = \mathsf{NecNot}$ then: $\mathsf{PrfNotAcc}^{\cap}(B, \mathsf{S}) = \emptyset$. Thus, there are no arguments necessary for the skeptical non-acceptance of B.

Note that these correspond to the intuitions in Example 39.

The next example shows how necessary and sufficient explanations differ from explanations based on extension-based semantics as discussed in Sections 3 and 4.

Example 41. Recall the argumentation framework $\mathcal{AF}_2$ from Examples 7 and 33. Recall that $\mathsf{Prf}(\mathcal{AF}_2) = \{\mathcal{E}_1, \mathcal{E}_2\} = \{\{A, C, D\}, \{A, C, E\}\}$.

- If $\mathbb{D} = \mathsf{SuffNot}$, then $\mathsf{PrfNotAcc}^{\cap}(B, \mathsf{S}) = \{\{C\}, \{C, D\}, \{D\}\}$ for $\mathsf{S} \in \{\mathcal{E}_1, \mathcal{E}_2\}$.
- If $\mathbb{D} = \mathsf{NecNot}$, then $\mathsf{PrfNotAcc}^{\cap}(B, \mathsf{S}) = \emptyset$ for $\mathsf{S} \in \{\mathcal{E}_1, \mathcal{E}_2\}$.

Thus, similar to Example 33 even though C is always part of the extension under any completeness-based semantics, $\{D\}$ is also sufficient for the non-acceptance of B. Therefore, C is not necessary for the non-acceptance of B either, since the acceptance of D is already sufficient.

6.3.2. Necessity and sufficiency for non-accepted formulas

Similar to the necessary and sufficient explanation for accepted formulas, the non-acceptance explanations for formulas might differ from the argument explanations if the elements of the explanations are taken into account. We therefore refine the non-acceptance explanation definitions as well, similar to Definition 30. Like before, we assume that $\mathsf{AllArgs}(\phi) \neq \emptyset$.

Definition 35. Let $\mathcal{AF}(\mathrm{AT}) = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework based on the argumentation theory AT and let $\phi \in \mathcal{L}$ be such that ϕ is not $\star$-accepted for $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$. Moreover, let $A_\phi \in \mathsf{AllArgs}(\phi)$ and $\mathsf{S} \in \mathsf{SuffNot}(A_\phi, \emptyset)$. First, the basic sufficient explanations, based on Definition 23:

$$\mathsf{SuffNotAcc}^{\cap}(\phi, \mathsf{S}) = \{\langle \mathbb{F}(A), \mathbb{F}(\mathsf{S}) \rangle \mid A \in \mathsf{AllArgs}(\phi), \mathsf{Conc}(A) = \phi \text{ and } \exists B \in \mathsf{S} \text{ such that } (B, A) \in \mathsf{Att}\}$$

$$\mathsf{SuffNotAcc}^{\cup}(\phi) = \{\langle \mathbb{F}(A), \mathbb{F}(\mathsf{S}) \rangle \mid A \in \mathsf{AllArgs}(\phi) \text{ and } \mathsf{S} \in \mathsf{SuffNot}(A, \emptyset)\}.$$

Now suppose that minimality has to be taken into account. Then based on the above and Definition 25, where $\preceq \in \{\leq, \subseteq\}$:

$$\mathsf{SuffNotAcc}^{\cap}_{\preceq}(\phi, \emptyset) = \left\{ \langle \mathbb{F}(A), \mathsf{S}' \rangle \;\middle|\; \mathsf{S}' \in \min^{\preceq}_{A \in \mathsf{AllArgs}(\phi) \& \mathsf{S} \in \mathsf{SuffNot}(A, \emptyset)} \mathbb{F}(\mathsf{S}) \right\}$$

$$\mathsf{SuffNotAcc}^{\cup}_{\preceq}(\phi) = \left\{ \left\langle \mathbb{F}(A), \mathbb{F}(\mathsf{S}') \right\rangle \;\middle|\; A \in \mathsf{AllArgs}(\phi) \text{ and } \mathsf{S}' \in \min^{\preceq}_{\mathsf{S} \in \mathsf{SuffNot}(A, \emptyset)} \mathbb{F}(\mathsf{S}) \right\}.$$

Finally, we consider necessity, based on a similar construction as the necessity explanations from Definition 30:

$$\mathsf{NecNotAcc}^{\cap}(\phi, \emptyset) = \left\{ \left\langle \mathbb{F}(A), \bigcap_{\mathsf{S} \in \mathsf{SuffNot}(A, \emptyset)} \mathbb{F}(\mathsf{S}) \right\rangle \;\middle|\; A \in \mathsf{AllArgs}(\phi) \right\}$$

$$\mathsf{NecNotAcc}^{\cup}(\phi) = \left\{ \left\langle \mathbb{F}(A), \bigcap_{\mathsf{S} \in \mathsf{SuffNot}(A, \emptyset)} \mathbb{F}(\mathsf{S}) \right\rangle \;\middle|\; A \in \mathsf{AllArgs}(\phi) \right\}.$$

The skeptical sufficient explanation returns for each argument A with conclusion ϕ attacked by the given set of arguments S a pair consisting of the elements of A and the elements of S determined by $\mathbb{F}$ (why is S sufficient for the non-acceptance of arguments for ϕ?). Thus, the explanations show why S is sufficient for the non-acceptance of ϕ, by giving for each argument A for ϕ in S a pair, consisting of A as a reason for why ϕ could be a conclusion and S in terms of (the elements of) the arguments. The credulous sufficient explanation is similar, but more extensive. It considers all arguments for ϕ and all sufficient sets for the non-acceptance of these arguments (what are the sufficient sets for the non-acceptance of all arguments for ϕ?). Thus, it returns the sufficient reasons for the non-acceptance of ϕ, considering all arguments and all sufficient sets in terms of (the elements of) the arguments.

The second set of explanations is similar, but takes minimality into account. In the skeptical case, minimality is considered over all arguments with conclusion ϕ and the sufficient sets for the non-acceptance of these arguments (what is a minimal sufficient set for the non-acceptance of an argument for ϕ?). The explanation is then a pair, containing an argument for ϕ and a sufficient set for its non-acceptance in terms of (the elements of) the arguments, such that the second element is minimal. In the credulous case all arguments for ϕ are considered and minimality is taken over the sufficient sets for the non-acceptance of each of these arguments (what are the minimal sufficient sets for the non-acceptance of all arguments for ϕ?).

In the case of sufficiency, $\mathbb{D} = \mathsf{SuffNot}$ is applied to the explanations from Definition 24. For necessity the intersection of the elements of these sufficient sets based on the following definition is applied:

- $\mathsf{Nec}(A, \emptyset) = \bigcap_{S \in \mathsf{SuffNot}(A,\emptyset)} \mathbb{F}(S)$, which denotes the set of all elements (determined by $\mathbb{F}$) of the intersection of the sufficient sets for A.

Like in the acceptance case, recall Definition 30, the skeptical and credulous necessity explanations are similar. This is the case since necessity is a skeptical property. The explanations are again sets of paris, the first element is an argument A for ϕ and the second contains all the necessary arguments or elements there of for the non-acceptance of A (what elements are necessary for the non-acceptance of ϕ?).

The resulting explanations for the running example are similar to those for the acceptance explanations:

Example 42. Recall that, for the argumentation framework $\mathcal{AF}(\mathrm{AT}_1)$, A is the only argument for u and B is the only argument for $\neg n(d_1)$. We have the following explanations:

- For sufficiency, if $\mathbb{F} = \mathsf{id}$: $\mathsf{SuffNotAcc}^{\cap}(u) \in \{\langle A, \{B\}\rangle, \langle A, \{D, F\}\rangle, \langle A, \{B, D, F\}\rangle\}$ and $\mathsf{SuffNotAcc}^{\cap}(\neg n(d_1)) \in \{\langle B, \{C\}\rangle, \langle B, \{E\}\rangle, \langle B, \{C, E\}\rangle\}$, thus B on its own, D and F or all three together are sufficient for the non-acceptance of argument A for u while C, E or C and E together are sufficient for the non-acceptance of argument B for $\neg n(d_1)$.
- For sufficiency, if $\mathbb{F} = \mathsf{Prem}$: $\mathsf{SuffNotAcc}^{\cap}(u) = \langle \{w\}, \{v\}\rangle$ and $\mathsf{SuffNotAcc}^{\cap}(\neg n(d_1)) \in \{\langle \{v\}, \{r, s\}\rangle, \langle \{v\}, \{r, t\}\rangle, \langle \{v\}, \{r, s, t\}\rangle\}$, so the premise v is sufficient on its own for the non-acceptance of u, while for the non-acceptance of $\neg n(d_1)$, r and s, r and t or r, s and t together are sufficient.
- For necessity, $\mathsf{NecNotAcc}^{\cap}(u) = \langle A, \{D, F\}\rangle$ and $\mathsf{NecNotAcc}^{\cap}(\neg n(d_1)) = \langle B, \emptyset\rangle$ if we have that $\mathbb{F} = \mathsf{id}$ and if $\mathbb{F} = \mathsf{Prem}$: $\mathsf{NecNotAcc}^{\cap}(u) = \langle \{w\}, \{v\}\rangle$ and $\mathsf{NecNotAcc}^{\cap}(\neg n(d_1)) = \langle \{v\}, \{r\}\rangle$, so, when we look at arguments, D and F are necessary for the non-acceptance of u, there is no argument necessary for the non-acceptance of or $\neg n(d_1)$, but when looking at the premises, we have that v is necessary for the non-acceptance of u and r is necessary for the non-acceptance of r.

6.3.3. Properties of necessity and sufficiency for non-acceptance

The next propositions are the non-acceptance counterparts of Propositions 7 and 8. First some basic properties of sufficiency and necessity for non-acceptance, showing the conditions under which sufficient and necessary explanations for non-acceptance are empty.

Proposition 9. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an argumentation framework and let $A \in Args$ be non-accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ *and $\star \in \{\cap, \cup\}$. Then:*

- $\mathsf{SuffNot}(A, \emptyset) \neq \emptyset$;
- $\mathsf{NecNot}(A, \emptyset) = \emptyset$ *implies that there are at least two direct attackers of A.*

Recall that non-acceptance explanations for $\mathbb{D} = \mathsf{NoDefAgainst}$ are never empty (Proposition 5), since there has to be a reason (an attack) that causes the non-acceptance. This is reflected in the results above: there is always some sufficient set of arguments for the non-acceptance of an argument, but it can be that there are several reasons for the non-acceptance of an argument, so it might be that there are no necessary arguments.

Now we show how $\mathbb{D} = \mathsf{NoDefAgainst}$ is related to the here introduced notions of sufficiency and necessity for non-acceptance. In particular, for all extensions $\mathcal{E}$, the first item shows that the set of arguments obtained by $\mathsf{NoDefAgainst}(A, \mathcal{E})$ is a sufficient set for the non-acceptance for A and the second item shows that $\mathsf{NoDefAgainst}(A, \mathcal{E})$ contains the necessary arguments for the non-acceptance of A. As a result, the basic explanations in Section 3.2 defined based on $\mathsf{NoDefAgainst}$ are sufficient and contain the necessary arguments for the non-acceptance for A.

Proposition 10. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an argumentation framework and let $A \in Args$ be an argument that is not accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ *and $\star \in \{\cap, \cup\}$. Then:*

- *for all $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ such that $A \notin \mathcal{E}$,* $\mathsf{NoDefAgainst}(A, \mathcal{E}) \in \mathsf{SuffNot}(A, \emptyset)$;
- $\mathsf{NecNot}(A, \emptyset) \subseteq \bigcap_{\mathcal{E} \in \mathsf{SemWithout}(A)} \mathsf{NoDefAgainst}(A, \mathcal{E})$.

Remark 14. The above proposition shows that necessary and sufficient explanations for non-acceptance cannot be calculated with Algorithm 1. For sufficiency, the reason is similar to the acceptance case (recall Remark 13): not all sufficient sets are calculated by

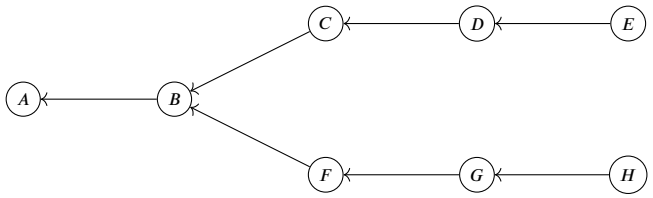


Fig. 8. Graphical representation of $\mathcal{AF}_8$.

the algorithm. For necessity, the algorithm can calculate $\mathsf{NoDefAgainst}(A, \mathcal{E})$ for all $\mathcal{E} \in \mathsf{SemWithout}(A)$ and it could therefore easily be adjusted to calculate their intersection, however, $\mathsf{NecNot}(A, \emptyset)$ might be a subset of the intersection.

6.4. Necessity, sufficiency and minimality

Although the notions of minimality from [33] and recalled in Section 6.1 are aimed at reducing the size of an explanation, by applying instead the notions of sufficiency and necessity as discussed in this paper, the size of the explanation can be further reduced. To see this, consider the following example:

Example 43. Let $\mathcal{AF}_8 = \langle \mathsf{Args}_8, \mathsf{Att}_8 \rangle$ as shown in Fig. 8. Here we have that $\mathsf{Prf}(\mathcal{AF}_8) = \{\mathcal{E}_8\} = \{\{A, C, E, F, H\}\}$. If $\mathbb{D}^{\mathsf{acc}} = \mathsf{Defending}$ and $\mathbb{D}^{\mathsf{nacc}} = \mathsf{NoDefAgainst}$, then $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_8) = \{C, E, F, H\}$ and $\mathsf{PrfNotAcc}^{\cap}(A, \mathcal{E}_8) = \{C, E, F, H\}$. These are the explanations for the acceptance of A and the non-acceptance of B as defined in Section 3 or Section 6.1.

Now consider $\mathbb{D}^{\mathsf{acc}} = \mathsf{MinSuff}^{\subseteq}$ and $\mathbb{D}^{\mathsf{nacc}} = \mathsf{MinSuffNot}^{\subseteq}$. Then $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_8) = \{\{A, C, E\}, \{A, F, H\}\}$ and $\mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_8) = \{\{C\}, \{F\}\}$. If $\mathbb{D}^{\mathsf{acc}} = \mathsf{Nec}$ and $\mathbb{D}^{\mathsf{nacc}} = \mathsf{NecNot}$ then $\mathsf{PrfAcc}^{\cup}(A, \mathcal{E}_8) = \{A\}$ and $\mathsf{PrfNotAcc}^{\cap}(B, \mathcal{E}_8) = \{C, E, F, H\}$. Indeed, $\{A, C, E\}$ or $\{A, F, H\}$ would already be sufficient for the acceptance of A, yet this cannot be expressed with minimal explanations. And, similarly, both $\{C\}$ and $\{F\}$ are minimally sufficient for the non-acceptance of B. Thus, the minimally sufficient explanations are $\leq$-smaller than the minimal explanations from Section 6.1.

That minimally sufficient explanations can be smaller than minimal explanations is formalized in the next propositions. We show that for any set in a $\cap$-acceptance explanation, there is a minimally sufficient explanation that can be smaller, that the reverse also holds for admissible semantics and that any explanation (whether minimal or not) contains all the necessary arguments.

Proposition 11. *Let $\mathcal{AF} = \langle Args, Att \rangle$ be an argumentation framework, let $A \in Args$ and let $\preceq \in \{\subseteq, \leq\}$. Then, for $\mathbb{D} = \mathsf{Defending}$ and $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$:*

- *for every $S \in \mathsf{SemAcc}^{\cap}(A)$ there is an $S' \in \mathsf{SemAcc}^{\cap}_{\preceq}(A)$ for $\mathbb{D} = \mathsf{MinSuff}^{\preceq}$ such that $S' \preceq S$;*
- *where $\mathsf{Sem} = \mathsf{Adm}$, for every $S \in \mathsf{SemAcc}^{\cap}_{\preceq}(A)$ where $\mathbb{D} = \mathsf{MinSuff}^{\preceq}$ also $S \in \mathsf{SemAcc}^{\cap}(A)$;*
- *for all $S \in \mathsf{SemAcc}^{\cap}(A)$, $\mathsf{Nec}(A, \emptyset) \subseteq S$.*

In words, for every (minimal) basic explanation, there is a minimal sufficient explanation that might be smaller (item 1), all minimal sufficient explanations are part of the $\cap$-explanation under admissibility semantics (item 2) and all (minimal) explanations still contain the necessary explanations (item 3).

The next proposition is the non-acceptance counterpart of the above proposition:

Proposition 12. *Let $\mathcal{AF} = \langle Args, Att \rangle$ be an argumentation framework, let $A \in Args$ and let $\preceq \in \{\leq, \subseteq\}$. Then, for $\mathbb{D} = \mathsf{NoDefAgainst}$ and $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Sstb}, \mathsf{Egr}\}$:*

- *for every $S \in \mathsf{SemNotAcc}^{\cup}(A)$ there is an $S' \in \mathsf{SuffNotAcc}^{\cup}_{\preceq}(A)$ where $\mathbb{D} = \mathsf{MinSuffNot}^{\preceq}$ such that $S' \preceq S$.*
- *for all $S \in \mathsf{SemNotAcc}^{\cup}(A)$, $\mathsf{NecNot}(A, \emptyset) \subseteq S$.*

Thus, for every (minimal) set of a $\cup$-non-acceptance explanation, there is a minimal sufficient non-acceptance explanation which might be smaller (item 1) and every set of a $\cup$-non-acceptance explanation (whether minimal or not) contains the necessary arguments for non-acceptance (item 2).

These propositions show that our implementation of necessity and sufficiency result in meaningful but smaller than minimal explanations. In the next section we will turn to a real-life application and illustrate the various explanations introduced in this paper.

7. Applying the basic framework

In this section we show how the various explanations can be applied in a real-life scenario. The Netherlands Police employs several applications based on structured argumentation frameworks (a variant of ASPIC$^+$, see [54] for the formal details[17]). One such application concerns complaints by citizens about online trade fraud (e.g., a product bought through a web-shop or on eBay never arrives or turns out to be fake). The system queries the citizen for various observations, and then determines whether the complaint is a case of fraud [12,54]. Another related example is a classifier for checking fraudulent web-shops, which scrapes information about online shops and then tries to determine whether they are real (bone fide) or fake (mala fide) shops [53]. These applications are aimed at assisting the police at working through high volume tasks, leaving more time for tasks that require human attention. In this paper we focus on applying the explanations in the context of the application on online trade fraud. This application is based on the formal modeling of Article 326 of the Dutch Criminal Code[18]:

> A person who takes some good or money away from someone, while misleading through false contact details, deceptive tricks or an accumulation of lies is guilty of fraud.

Consider a scenario where a complainant has ordered a product through an online shop. We introduce the following language $\mathcal{L}_2$: the complainant delivered (*cd*), the counterparty delivered (*cpd*); the received product seems fake (*fake*); the counterparty provided contact details (*con*); the contact details were false (*fcon*); the received package is indeed fake (*recfake*); the delivery may still arrive (*deco*); it is a case of fraud (f); and their negations. While the complainant provides the information from the described scenario (the knowledge base of this particular instance), the system constructs further arguments from this, based on the Dutch Criminal Law.[19] Important in Article 326 is that money was taken away from someone: the complainant (i.e., the supposed victim) delivered, that is, they kept their side of the bargain by paying, and the counterparty (i.e., the possible suspect) did not deliver, that is, they did not send the (right) product or package to the complainant. Also, simply not delivering is not considered fraud: the counterparty must have shown their intent to commit fraud by, for example, providing fake contact details so the complainant could not contact them about the (undelivered) package, or by, for example, trying to pass off a fake product as real. This leads to the following defeasible rules:

- R_1 If the counterparty delivered (*cpd*) and the received product seems fake (*fake*) then the received packages are usually indeed fake (*recfake*);
- R_2 If the counterparty did not deliver (¬*cpd*) and no contact details were provided (¬*con*) then usually the delivery will not arrive (¬*deco*);
- R_3 If the counterparty did not deliver (¬*cpd*), contact details were provided (*con*) and the contact details were false (*fcon*) then usually the delivery will not arrive anymore (¬*deco*);
- R_4 If the counterparty did not deliver (¬*cpd*), contact details were provided (*con*) and the contact details were not false (¬*fcon*) then usually the package may still arrive (*deco*);
- R_5 If the complainant delivered (*cd*) and the received package is indeed fake (*recfake*) it is usually a case of fraud (f);
- R_6 If the counterparty did not deliver (¬*cpd*), the complainant delivered (*cd*) and the delivery will not arrive anymore (¬*deco*) it is usually a case of fraud (f);
- R_7 If the delivery may still arrive (*deco*) and the complainant delivered (*cd*) it is usually not a case of fraud (¬f);
- R_8 If the product does not seem fake (¬*fake*) and the complainant delivered (*cd*) it is usually not a case of fraud (¬f);
- R_9 If the complainant did not deliver (¬*cd*) it is usually not a case of fraud (¬f).

Suppose we have the full knowledge base containing all the antecedents of rules that are not a consequent of a rule as well: the counterparty did (not) deliver (*cpd*, ¬*cpd*), the received product does (not) seem fake (*fake*, ¬*fake*), contact details were (not) provided (*con*, ¬*con*), the contact details were (not) false (*fcon*, ¬*fcon*), the complainant did (not) deliver (*cd*, ¬*cd*). We can then derive the arguments $A_1, \ldots, A_{10}$, $B_1, \ldots, B_4$ and $C_1, \ldots, C_6$ as shown in Fig. 9 on page 33. A graphical representation of the resulting argumentation framework $\mathcal{AF}_9$ is shown in Fig. 10 in Appendix B, on page 44. Since there are over 30 preferred extensions, we provide in the same appendix an exhaustive list of the preferred extensions, the argument explanations and the derivation process.

By design, the application tries to determine whether the provided scenario by the complainant is a case of fraud (f is accepted) or not (¬f is accepted). For both conclusions there are three arguments: C_1, C_2 and C_3, respectively C_4, C_5 and C_6. There are 10 preferred extensions with an argument for f and 25 preferred extensions with an argument for ¬f. As a result, simply returning the extensions containing an argument for f or ¬f is not very informative as an explanation of why something is (not) a case of fraud, especially for the non-expert citizen. But, returning the basic explanations as given in Section 3, also results in 10 possible explanations for arguments with conclusion f and 26 possible explanations for arguments with conclusion ¬f (see Appendix B for a complete overview of all extensions and all types of explanations). Each of these basic explanations contains a number of arguments

[17] The corresponding demo of [54], demonstrating the argumentation-based part of the application, is available at https://pyarg.npai.science.uu.nl, as example application.

[18] The article can be found, in Dutch, at: https://wetten.overheid.nl/BWBR0001854/2018-01-01/#BoekTweede_TiteldeelXXV_Artikel326.

[19] In order to make the argumentation framework and corresponding explanations more interesting the rules that are applied here are only inspired by the law. The real application is based on slightly different rules.

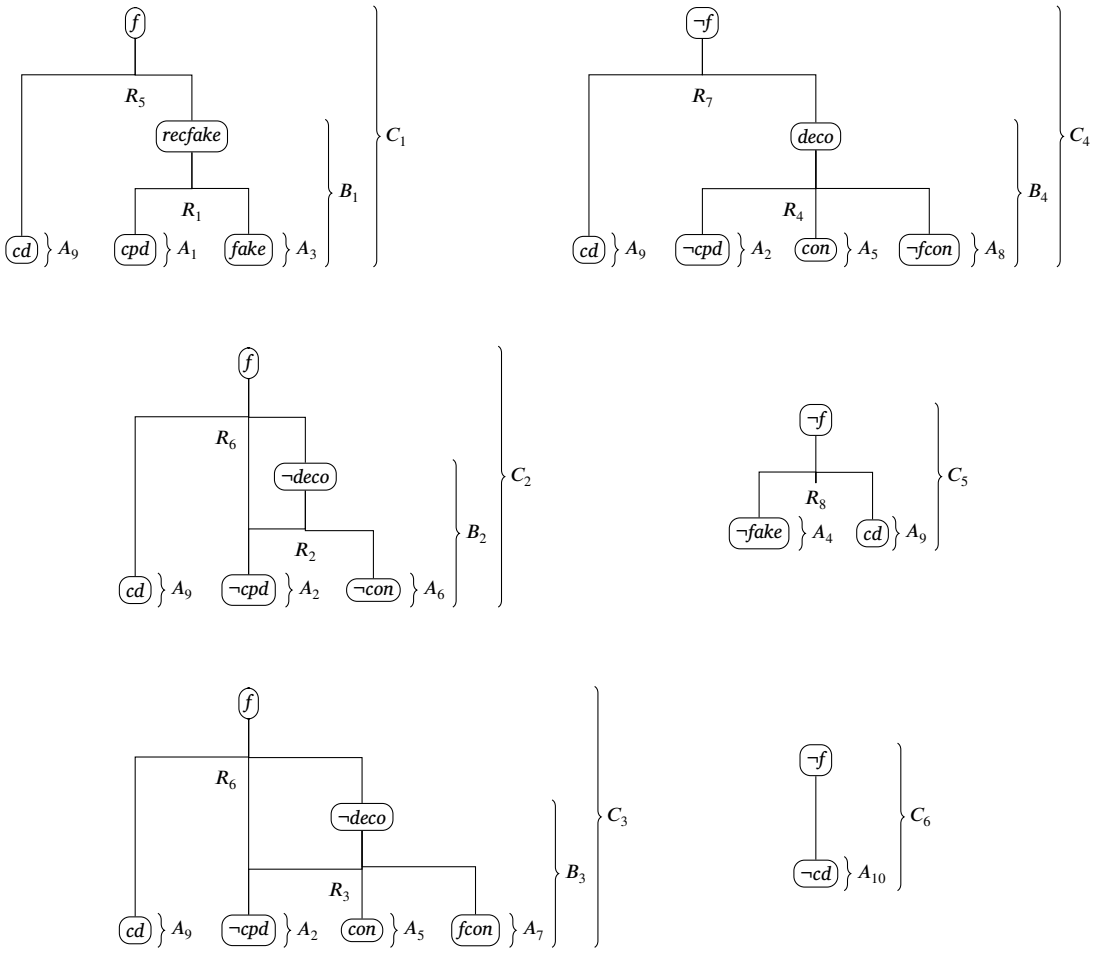


Fig. 9. Graphical representation of the derivations of the arguments in $\mathcal{AF}_9$.

which are constructed from a number of premises. Reducing the size of the explanations is therefore essential. Below we are therefore interested in minimal acceptance explanations denoted by $\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\preceq}(f,\mathcal{E})$ and $\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\preceq}(\neg f,\mathcal{E})$ for $\preceq \in \{\leq, \subseteq\}$ and $\mathcal{E} \in \{\mathcal{E}_1,\ldots,\mathcal{E}_{35}\}$. The full explanations are provided in Appendix B.

One other aspect of explanations that we can vary in this example is $\mathbb{F}$. For example, we can give back the arguments, premises, rules, and so on. Exactly how we choose $\mathbb{F}$ depends on the receiver of the explanation. For a police analyst, for example, it makes sense to pick $\mathbb{F} = \mathsf{id}$, as they will probably know what the arguments stand for (e.g., which conclusions they have) so they can work with a 'summary' in terms of argument id's. However, a citizen making a one-time complaint is probably less interested (and less familiar) with the specific arguments and their interactions. In the intake system at the police, citizens only provide basic information (i.e., premises of arguments) to the system, and it therefore makes sense to provide explanations in terms of the same kind of basic information provided by the citizens. We therefore choose $\mathbb{F} = \mathsf{Prem}$ and obtain the following minimal explanations for f and $\neg f$:

$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\leq}(f,\mathcal{E}_{33}) = \langle\{\neg cpd,\neg con,cd\},\{\neg cpd,\neg con,cd\}\rangle$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\subseteq}(f,\mathcal{E}_{3}) = \{\langle\{cpd,fake,cd\},\{cpd,fake,con,cd\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\subseteq}(f,\mathcal{E}_{7}) = \{\langle\{cpd,fake,cd\},\{cpd,fake,\neg con,cd\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\subseteq}(f,\mathcal{E}_{23}) = \{\langle\{\neg cpd,\neg con,cd\},\{\neg cpd,fake,\neg con,cd\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\subseteq}(f,\mathcal{E}_{25}) = \{\langle\{\neg cpd,con,fcon,cd\},\{\neg cpd,con,fcon,cd\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\subseteq}(f,\mathcal{E}_{33}) = \langle\{\neg cpd,\neg con,cd\},\{\neg cpd,\neg con,cd\}\rangle$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\leq}(\neg f,\mathcal{E}_{2}) = \{\langle\{\neg cd\},\{cpd,con,\neg cd\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\leq}(\neg f,\mathcal{E}_{6}) = \{\langle\{\neg cd\},\{cpd,\neg con,\neg cd\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}_{\leq}(\neg f,\mathcal{E}_{18}) = \{\langle\{\neg cd\},\{\neg cpd,con,\neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{22}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{2}) = \{\langle\{\neg cd\}, \{cpd, con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{6}) = \{\langle\{\neg cd\}, \{cpd, \neg con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{9}) = \{\langle\{\neg fake, cd\}, \{cpd, \neg fake, con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{13}) = \{\langle\{\neg fake, cd\}, \{cpd, \neg fake, \neg con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{18}) = \{\langle\{\neg cd\}, \{\neg cpd, con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{19}) = \{\langle\{\neg cpd, con, \neg fcon, cd\}, \{\neg cpd, con, \neg fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{22}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{26}) = \{\langle\{\neg fake, cd\}, \{\neg cpd, \neg fake, con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{31}) = \{\langle\{\neg fake, cd\}, \{\neg cpd, \neg fake, \neg con, cd\}\rangle\}.$$

These explanations suggest that there is a variety of scenarios possible in which there is a case of fraud (or not). For example, it is a case of fraud when the counterparty did not deliver ($\neg cpd$), no contact details were provided by the counterparty ($\neg con$) and the complainant delivered (cd), or when both the complainant and the counterparty delivered (cd and cpd), contact details were provided (con) but the received product seems fake ($fake$). Similarly, it is not a case of fraud when both the complainant and the counterparty delivered, contact details were provided and the received product seems not to be a fake.

The above example shows that (subset-)minimality can be a rather crude measure for explanations. Even for a relatively simple structured argumentation framework, the number of $\subseteq$-minimal explanations is quite large, and $\leq$-minimality misses certain explanations (e.g., about fake products) because these arguments have more premises. For more meaningful and smaller explanations we turn to necessity and sufficiency, obtaining the following explanations for $\leq$-minimal sufficiency and necessity:

$$\text{Minimal sufficiency:}\quad \mathsf{Suff\,Acc}^{\cup}_{\leq}(f, \emptyset) = \{\langle\{cpd, fake, cd\}, \{cpd, fake, cd\}\rangle, \langle\{\neg cpd, \neg con, cd\}, \{\neg cpd, \neg con, cd\}\rangle\}$$
$$\text{Necessity:}\quad \mathsf{NecAcc}^{\cup}(f, \emptyset) = \langle\{cd\}, \{cd\}\rangle$$
$$\text{Minimal sufficiency:}\quad \mathsf{Suff\,Acc}^{\cup}_{\leq}(\neg f, \emptyset) = \{\langle\{\neg cd\}, \{\neg cd\}\rangle\}$$
$$\text{Necessity:}\quad \mathsf{NecAcc}^{\cup}(\neg f, \emptyset) = \emptyset.$$

As mentioned above, there is a variety of possible scenarios for both fraud and not fraud, but based on the necessity and sufficiency explanations one knowledge base element is essential: the complainant must have delivered for fraud. With the basic explanations, even when only looking at the minimal ones, this conclusion could not be derived and the explanation might have been $\{cpd, \neg con, \neg cd\}$ for not fraud, while the explanation only containing $\neg cd$ would be more informative. Therefore, with necessary and sufficient explanations, we can provide compact explanations that only contain the core reasons for a conclusion: it allows to point to a specific part of the law that resulted in this particular conclusion. This was not possible with the (minimal) explanations from the basic framework.

8. Related work

Explainable AI is a fast growing research area and explainability is investigated in many sub-areas of AI, including in formal argumentation. As mentioned in the introduction, one of the main contributions of this paper is our focus on necessity and sufficiency. These are some of the most important selection mechanisms of explanations [46] and have been studied in logic, philosophy and psychology [37,41–43,66]. Necessity and sufficiency has been implemented in some approaches to XAI as well. For example, [65] propose a unifying framework, where necessity and sufficiency are based on probabilities and in [28] pertinent positives and pertinent negatives are studied. Although these approaches help to select the most relevant explanations, these do not increase the transparency of the black-box models on which they are applied. In this paper we focused on necessity and sufficiency for argumentation.

When it comes to literature on argumentation-based explanations, there are roughly two directions. First, there is the application of formal argumentation to explain decisions derived with other AI-methods, e.g., [21,26,59], see [24] for a recent overview. These approaches can be seen as a step towards hybrid, or neuro-symbolic AI, where the performance of learning-based methods is combined with the transparency of knowledge-based methods (in this case formal argumentation) [25,44,45]. Second, there is literature on local explanations of formal argumentation, where argumentation is applied to explain argumentation-based conclusions, see e.g., [19,27,33,34,36,40,52,61]. For now, our research falls within this second direction of research and we will relate this paper to the existing literature below. As we will mention in the next section, with this research we aim to develop an argumentation-based explanations method that is generally applicable and can improve the argumentation-based component of the research in the first direction. Therefore, our research is also relevant for the first direction, though not directly related yet.

García et al. [36] study explanations for abstract argumentation and DeLP [35]. Explanations for a claim are defined as triples of dialectical trees that provide a warrant for the claim, dialectical trees that provide a warrant for the contrary of the claim, and dialectical trees for the claim and its contrary that provide no warrant. This means, on the one hand, that explanations might contain many arguments and, on the other hand, that the receiver of the explanation is expected to understand argumentation and dialectical

trees. With real-life applications in mind, we believe that explanations that rely less on the underlying argumentation framework and that can be adjusted to the application would be more useful.

To explain the acceptance of an argument in abstract argumentation and ABA [15], the notion of *related admissibility* is introduced in [33] and explanations are derived based on dispute trees [31]. In [16, Proposition 2], it is shown that these explanations are a special case of our basic argument explanations. These explanations based on dispute trees are applied in, e.g., [23,67]. Our explanations framework is more general, for example, we take the semantics into account and allow to refine the explanation based on the underlying structure of the arguments, if this is available. The method in [19] is also based on a (formal) discussion, but rather than investigating the explanatory properties of these discussions, that paper is on the computational complexity.

Explanations for non-accepted arguments in abstract argumentation are studied in [34,61], both of which focus on the structure of the AF and not credulous acceptance under admissible semantics. Note that we consider a skeptical and credulous variant of non-acceptance for several Dung-style semantics. In [34] an explanation consists of either a set of arguments or a set of attacks, the removal of which would make the argument admissible. In structured argumentation it is not always possible to remove exactly one argument (or attack). In $\mathcal{AF}_9$ of Fig. 10 from Section 7, A_1 would become skeptically accepted for any completeness-based semantics, if A_2 would be removed. However, when looking at the underlying argumentation theory, when A_2 is removed, the arguments B_2 and B_3 do no longer exist and thus $\neg deco$ is no longer a credulous conclusion. Therefore, in this paper the basic definition for non-accepted arguments is defined in terms of the arguments for which no defense exist and no suggestion is made how to change the argumentation framework in order to get the considered argument accepted. In [61], explanations are sub-frameworks, such that the considered argument is not credulously accepted in that sub-framework and any of its super-frameworks. Though a note was added on the applicability of such explanations in a structured setting, this is not formally investigated in that paper. Computational complexity for the explanations in [61] is studied in [52], but still only for non-accepted arguments in abstract argumentation.

Like Saribatur et al. [61], Ulbricht and Wallner [63] use the notion of subframeworks to study strong explanations: sets of arguments that ensure that, when present in a subframework, the desired set of arguments is acceptable under admissibility in that subframework. It is shown that these explanations result in smaller sets of arguments than the extensions, sufficient for the acceptability. Since this work relies on the structure of the argumentation framework, it is designed for abstract argumentation however, as illustrated above, in structured argumentation it is not possible to consider subframeworks without considering the underlying structure of the arguments. Our work is applicable to both abstract and structured settings, for both acceptance and non-acceptance and we consider both sufficiency and necessity.

In [40] explanations are presented as a semantics that assigns to each accepted argument a set of explanations arguments. Intuitively, the arguments in the explanation are together sufficient to accept the argument. The properties of such explanations are then studied with a principle-based approach. Finally, while in all the previous mentioned literature extension-based semantics is applied, in [27] the interpretability of abstract argumentation with gradual semantics is studied.

Since the introduction of our notions of necessity and sufficiency in [17], similar notions have been studied. For example, Potyka et al. [55] use these notions, combined with bipolar argumentation and Markov networks to explain random forests. Rotolo and Sartor [60] study the relation between argumentation and explanation in the law in a formal setting. While their notions of necessity and sufficiency are similar, they focus on specific legal conclusions rather than general arguments or their elements. Besnard et al. [10] provide visual explanations in the context of abstract argumentation and although they do write about sufficient reasons, this notion is not formally defined. Finally, Kampik et al. [38] also provide sufficient and necessary explanations, but their definitions differ since their explanations are focused on change in Quantitative Bipolar Argumentation Frameworks (QBAFs) [6].

Proof theories and argument games are related to argumentation-based explanations. For example, in [29], a proof theory for skeptical acceptance under preferred semantics is introduced, answering the question why an argument is skeptically accepted and in [47] argument games for several extension-based semantics are collected. Although the explanations from Section 3 can be represented in terms of argument games, the variations to these basic explanations and in particular the notions of sufficiency and necessity have not been studied. Moreover, our work is designed to be applicable to structured settings as well. Acceptance for formulas might differ from acceptance for arguments (e.g., in $\mathcal{AF}_6$ from Example 25 p is skeptically accepted, but no argument for p is skeptically accepted) and, as a result, formula explanations might differ from argument explanations. To account for this difference is essential, since the case study from Section 7 shows that there are real-life applications based on structured settings that might benefit from our explanations. Representing explanations as an argument or discussion game is of interest when preparing explanations for applications and in future work we will study how the variations to the basic explanations can be represented.

To the best of our knowledge, this is the only framework that can explain acceptance and non-acceptance in a similar way and that is generally applicable: it does not depend on the specific structure of an argumentation framework, the applied extension-based semantics or the way in which the argumentation framework was derived (i.e., it is applicable to both abstract and structured argumentation). Additionally, we have ensured that our framework is easily adjustable to the application and receiver at hand, by introducing the functions $\mathbb{D}$ and $\mathbb{F}$. As a result, our framework is more general than the existing approaches to local argumentation-based explanations.

9. Conclusion and future work

Argumentation has been gaining traction as a valid AI technique for practical applications [3,53,54], and is also increasingly used to explain decisions derived with other AI-approaches [24,64]. With formal argumentation frameworks thus being increasingly used and becoming more complex, it is essential that we can explain the conclusions of such an argumentation framework (i.e., which arguments are accepted and which ones are not) in terms of *relevant arguments*, and that we can select the explanations that conform

to criteria for 'good' explanations (cf. [46]). In this paper, we have developed an explanation method that does exactly this, and is generally applicable and adjustable to both abstract and structured argumentation frameworks. In particular, we have made the following contributions:

- We have proposed a generic, flexible formal framework for determining and computing argument-based explanations for arguments and their conclusions (Sections 3 and 5). The framework can be applied in abstract and structured settings, for accepted and non-accepted arguments and formulas and for a variety of Dung-style semantics. It can provide different types of useful explanations, in terms of arguments and formulas.
- We have shown the flexibility of the framework by implementing how humans select relevant explanations based on minimality, necessity and sufficiency (Section 6).
- In Section 7 we have applied the framework to a real-life scenario of the Netherlands Police. This demonstrates the usefulness of our framework as well as the benefits of necessity and sufficiency.

Together, this shows the potential of our framework to be applied as a general argumentation-based explanations method. Additionally, our study on minimality, necessity and sufficiency is of interest beyond explanations.

- We have shown how the essential information can be selected from extensions and argumentation settings in general. Therefore, our work paves the way to investigate how to draw conclusions from formal argumentation beyond the well-known extensions.

With the presented framework, we will work towards an explanation method that any researcher in explainable AI can apply to explain decisions from their AI-application in a reliable and human understandable way. Therefore, in future work, we will study which findings from the social sciences [46] (e.g., contrastiveness and other selection methods) and the argumentation literature (e.g., argument-based dialogues [13]) should be accounted for. We will also continue our investigation into the implementation of explanations in the real-life applications at the Netherlands Police. To this end we will do a user study with both citizens and police analysts as well as improve the computational complexity of our method.

In addition to being interesting criteria for selecting an explanation, minimality, necessity and sufficiency are also of interest when studying what information (e.g., arguments or formulas) makes it possible or ensures that an argument or claim is accepted in formal argumentation in general. Thus, our findings of this paper will also be interesting for studying properties of semantics as well as in dynamic settings. In dynamic argumentation, the argumentation framework can be adjusted in view of new information, reflecting the dynamic nature of argumentation. In enforcement [7,18], for example, the question is whether it is possible to adjust the framework such that an argument or formula becomes accepted. Knowing which sets of arguments are necessary or sufficient for the (non-)acceptance of an argument is therefore useful.

CRediT authorship contribution statement

AnneMarie Borg: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Floris Bex:** Conceptualization, Formal analysis, Funding acquisition, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This research has been partly funded by the Dutch Ministry of Justice and the Netherlands Police.

Appendix A. Full proofs of the results in the paper

Throughout the paper no proofs were provided for our results. In this appendix full proofs of all the results are collected. For the purpose and intuition of the results we refer to their first statement within the paper itself.

A.1. Proofs of the properties of the basic explanations

Proposition 1. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ *be an argumentation framework where* $A \in \mathit{Args}$*. Then:*

1. $\mathsf{GrdAcc}(A) \subseteq \bigcap \mathsf{SemAcc}^{\cap}(A)$ *for all* $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$.

2. $\mathsf{StbAcc}^{\cap}(A) \subseteq \mathsf{SstbAcc}^{\cap}(A) \subseteq \mathsf{PrfAcc}^{\cap}(A) \subseteq \mathsf{CmpAcc}^{\cap}(A)$.
3. *For each* $\mathsf{S} \in \mathsf{CmpAcc}^{\cap}(A)$ *there is an* $\mathsf{S}' \in \mathsf{PrfAcc}^{\cap}(A)$ *such that* $\mathsf{S} \subseteq \mathsf{S}'$

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and let $A \in \mathsf{Args}$ be accepted.

1. Let $B \in \mathsf{GrdAcc}(A)$, then B (in)directly defends A and $B \in \mathsf{Grd}(\mathcal{AF})$, therefore $B \in \bigcap \mathsf{Sem}(\mathcal{AF})$. It follows that $B \in \mathsf{Defending}(A, \mathcal{E})$ for all $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$. Hence $B \in \bigcap \mathsf{SemAcc}^{\cap}(A)$ and $B \in \mathsf{SemAcc}^{\cup}(A)$ as well.
2. Let $\mathsf{S} \in \mathsf{StbAcc}^{\cap}(A)$, then there is some $\mathcal{E} \in \mathsf{Stb}(\mathcal{AF})$ such that $\mathsf{S} = \mathsf{Defending}(A, \mathcal{E})$. Since $\mathsf{Stb}(\mathcal{AF}) \subseteq \mathsf{Sstb}(\mathcal{AF}) \subseteq \mathsf{Prf}(\mathcal{AF}) \subseteq \mathsf{Cmp}(\mathcal{AF})$ it follows that $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for $\mathsf{Sem} \in \{\mathsf{Sstb}, \mathsf{Prf}, \mathsf{Cmp}\}$ as well. Therefore $\mathsf{S} \in \mathsf{SemAcc}^{\cap}(A)$ for $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Prf}, \mathsf{Sstb}\}$.
3. Let $\mathsf{S} \in \mathsf{CmpAcc}^{\cap}(A)$, then there is some $\mathcal{E} \in \mathsf{Cmp}(\mathcal{AF})$ such that $\mathsf{S} = \mathsf{Defending}(A, \mathcal{E})$. By Definition 1 there is some $\mathcal{E} \subseteq \mathcal{E}' \in \mathsf{Prf}(\mathcal{AF})$, let $\mathsf{S}' = \mathsf{Defending}(A, \mathcal{E}')$. Note that $\mathsf{S}' \in \mathsf{PrfAcc}^{\cap}(A)$ and that $\mathsf{S} \subseteq \mathsf{S}'$. □

Proposition 2. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ *be an argumentation framework,* $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ *for* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ *and let* $A, B \in \mathit{Args}$. *Then:*

- *if* $A \in \mathsf{Defending}(B, \mathcal{E})$, *then* $\mathsf{Defending}(A, \mathcal{E}) \subseteq \mathsf{Defending}(B, \mathcal{E})$;
- *if* $A \in \mathsf{Defending}(B, \mathcal{E})$ *and* $B \in \mathsf{Defending}(A, \mathcal{E})$, *then* $\mathsf{Defending}(A, \mathcal{E}) = \mathsf{Defending}(B, \mathcal{E})$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and let $A, B \in \mathsf{Args}$. Suppose that $A \in \mathsf{Defending}(B, \mathcal{E})$. By definition of Defending it follows that $A \in \mathcal{E}$. Let $C \in \mathsf{Args}$ such that $C \in \mathsf{Defending}(A, \mathcal{E})$. Then there is some $D \in \mathsf{Args}$ such that $(D, A) \in \mathsf{Att}$ and C defends A against this attack. However, since A defends B, it follows that D attacks B as well, from which it follows that C defends B as well. Therefore $C \in \mathsf{Defending}(B, \mathcal{E})$. The second item follows immediately. □

Proposition 3. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ *be an argumentation framework and let* $A \in \mathit{Args}$ *be such that* A *is accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ *and* $\star \in \{\cap, \cup\}$. *Then* $\mathsf{SemAcc}^{\cap}(A) = \{\emptyset\}$ *and* $\mathsf{SemAcc}^{\cup}(A, \mathcal{E}) = \emptyset$ *for any* $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ *iff there is no* $B \in \mathit{Args}$ *such that* $(B, A) \in \mathit{Att}$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A \in \mathsf{Args}$ be such that A is accepted w.r.t. $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ and $\star \in \{\cap, \cup\}$.

⇒ Suppose that $\mathsf{SemAcc}^{\cap}(A, \mathcal{E}) = \{\emptyset\}$ for all $\mathcal{E} \in \mathsf{SemWith}(A)$. Then for each $\mathcal{E} \in \mathsf{SemWith}(A)$, $\mathsf{Defending}(A, \mathcal{E}) = \emptyset$. Hence there is no attacker of A that is defended by some argument from $\mathcal{E}$. Since $A \in \mathcal{E}$, A is defended against its attackers. It follows that A is not attacked at all.

Similarly, suppose that $\mathsf{SemAcc}^{\cup}(A) = \emptyset$. Then there is some $\mathcal{E} \in \mathsf{SemWith}(A)$ such that $\mathsf{Defending}(A, \mathcal{E}) = \emptyset$. Hence there is no attacker of A that is defended by some argument from $\mathcal{E}$. Since $A \in \mathcal{E}$, A is defended against its attackers. It follows that A is not attacked at all.

⇐ Now suppose that A is not attacked. Then there is no argument that defends A. Therefore, for any $\mathcal{E} \in \mathsf{SemWith}(A)$, $\mathsf{Defending}(A, \mathcal{E}) = \emptyset$. It follows that $\mathsf{SemAcc}^{\cap}(A) = \{\emptyset\}$ and $\mathsf{SemAcc}^{\cup}(A, \mathcal{E}) = \emptyset$ for any $\mathcal{E}$. □

Proposition 4. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ *be an argumentation framework where* $A \in \mathit{Args}$. *Then:*

1. $\mathsf{GrdNotAcc}(A) \subseteq \bigcap \mathsf{CmpNotAcc}^{\cup}(A)$.
2. $\mathsf{SstbNotAcc}^{\cup}(A) \subseteq \mathsf{PrfNotAcc}^{\cup}(A) \subseteq \mathsf{CmpNotAcc}^{\cup}(A)$.
3. *Let* $\mathcal{E} \in \mathsf{SstbWithout}(A)$. *Then* $\mathsf{SstbNotAcc}^{\cap}(A, \mathcal{E}) = \mathsf{SemNotAcc}^{\cap}(A, \mathcal{E})$, *for* $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Prf}\}$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and let $A \in \mathsf{Args}$ be non-accepted.

1. Let $\mathcal{E}_{\mathsf{grd}} \in \mathsf{Grd}(\mathcal{AF})$ be the grounded extension, note that we have $\mathcal{E}_{\mathsf{grd}} = \bigcap \mathsf{Cmp}(\mathcal{AF})$ and that $\mathcal{E}_{\mathsf{grd}} \in \mathsf{Cmp}(\mathcal{AF})$ [30]. It therefore follows that, if $A \notin \mathcal{E}_{\mathsf{grd}}$, then $\mathcal{E}_{\mathsf{grd}} \in \mathsf{CmpWithout}(A)$. By Definition 8 we have that $\mathsf{GrdNotAcc}(A) \in \mathsf{CmpNotAcc}^{\cup}(A)$ and since $\mathcal{E}_{\mathsf{grd}} = \bigcap \mathsf{Cmp}(\mathcal{AF})$, also $\mathsf{GrdNotAcc}(A) \subseteq \bigcap \mathsf{CmpNotAcc}^{\cap}(A)$.
2. Suppose that A is not accepted w.r.t. Sstb, then for some $\mathcal{E} \in \mathsf{Sstb}(\mathcal{AF})$, $A \notin \mathcal{E}$. Note that $\mathsf{SstbWithout}(A) \subseteq \mathsf{PrfWithout}(A) \subseteq \mathsf{CmpWithout}(A)$. Therefore, for all $\mathcal{E} \in \mathsf{SstbWithout}(A)$, $\mathcal{E} \in \mathsf{PrfWithout}(A)$ and $\mathcal{E} \in \mathsf{CmpWithout}(A)$. Hence it holds that $\mathsf{SstbNotAcc}^{\cup}(A) \subseteq \mathsf{PrfNotAcc}^{\cup}(A) \subseteq \mathsf{CmpNotAcc}^{\cup}(A)$.
3. Let $\mathcal{E} \in \mathsf{SstbWithout}(A)$. As mentioned above, $\mathcal{E} \in \mathsf{PrfWithout}(A)$ and $\mathcal{E} \in \mathsf{CmpWithout}(A)$ as well. It follows that, $\mathsf{SstbNotAcc}^{\cap}(A, \mathcal{E}) = \mathsf{NoDefAgainst}(A, \mathcal{E}) = \mathsf{CmpNotAcc}^{\cap}(A) = \mathsf{PrfNotAcc}^{\cap}(A)$. □

Proposition 5. *Let* $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ *be an argumentation framework and let* $A \in \mathit{Args}$ *be such that* A *is non-accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Sstb}, \mathsf{Egr}\}$ *and* $\star \in \{\cap, \cup\}$. *Then* $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) \neq \emptyset$ *for any* $\mathcal{E} \in \mathsf{SemWithout}(A)$ *and* $\mathsf{SemNotAcc}^{\cup}(A) \neq \{\emptyset\}$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ and $A \in \mathsf{Args}$ be such that A is non-accepted w.r.t. $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$. Assume, towards a contradiction, that $\mathsf{SemNotAcc}^{\cup}(A) = \{\emptyset\}$, then for each extension $\mathcal{E} \in \mathsf{SemWithout}(A)$, $\mathsf{NoDefAgainst}(A, \mathcal{E}) = \emptyset$. Hence there is no $B \in \mathsf{Args}$ such that $(B, A) \in \mathsf{Att}$. But then, by the completeness of $\mathcal{E}$ it follows that $A \in \mathcal{E}$. A contradiction. Now assume that $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) = \emptyset$ for some $\mathcal{E} \in \mathsf{SemWithout}(A)$, then $\mathsf{NoDefAgainst}(A, \mathcal{E}) = \emptyset$. But then, as before, $A \in \mathcal{E}$. A contradiction. Therefore $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) \neq \emptyset$ for any $\mathcal{E} \in \mathsf{SemWithout}(A)$ and $\mathsf{SemNotAcc}^{\cup}(A) \neq \{\emptyset\}$. □

Proposition 6. *Let $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ be an argumentation framework, let $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for some $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and let $A, B_1, \ldots, B_n, C_1, \ldots, C_k \in \mathit{Args}$ such that $(B_1, A), \ldots, (B_n, A) \in \mathit{Att}$ and A indirectly attacks $C_1, \ldots, C_k$. Then:*

- *where $B_1, \ldots, B_m \in \mathcal{E}$, for $m \leq n$ it holds that:* $\mathsf{NoDefAgainst}(A, \mathcal{E}) \supseteq \mathsf{Defending}(B_1, \mathcal{E}) \cup \ldots \cup \mathsf{Defending}(B_m, \mathcal{E})$;
- *when $A \in \mathcal{E}$ we have:* $\mathsf{Defending}(A, \mathcal{E}) \subseteq \mathsf{NoDefAgainst}(B_1, \mathcal{E}) \cup \ldots \cup \mathsf{NoDefAgainst}(B_n, \mathcal{E})$;
- *where $A \in \mathcal{E}$ and $C_1, \ldots, C_j \notin \mathcal{E}$, for $j \leq k$ it holds that:* $\mathsf{Defending}(A, \mathcal{E}) \subseteq \mathsf{NoDefAgainst}(C_i, \mathcal{E})$ *for all* $i \in \{1, \ldots, j\}$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF, let $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for some $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and let $A \in \mathsf{Args}$.

- Let $B \in \mathcal{E}$ be such that $(B, A) \in \mathsf{Att}$. If $\mathsf{Defending}(B, \mathcal{E}) = \emptyset$, we are done, hence, let $C \in \mathsf{Defending}(B, \mathcal{E})$. Then, by the proof of Proposition 3 there is some $D \in \mathsf{Args}$ such that $(D, B) \in \mathsf{Att}$ and C (in)directly attacks D. Since B attacks A, it follows that D defends A and that C (in)directly attacks A. Since $C \in \mathcal{E}$, $\mathcal{E}$ does not defend A against the attack from C and therefore $C \in \mathsf{NoDefAgainst}(A, \mathcal{E})$.
- Let $C \in \mathsf{Defending}(A, \mathcal{E})$, then, by Proposition 3, $n \neq 0$. Suppose that C directly defends A, then there is a $B_i \in \{B_1, \ldots, B_n\}$ such that $(C, B_i) \in \mathsf{Att}$. Since $C \in \mathcal{E}$ it follows that $C \in \mathsf{NoDefAgainst}(B_i, \mathcal{E})$. Now suppose that C indirectly defends A. Then there are $D_1, D_2, \ldots, D_k \in \mathsf{Args}$, where k is odd, such that $(D_1, B_i), (D_2, D_1), \ldots (D_k, D_{k-1}), (C, D_k) \in \mathsf{Att}$. Since D_k defends B_i and C attacks D_k it follows that C attacks B_i as well. Hence $C \in \mathsf{NoDefAgainst}(B_i, \mathcal{E})$. Note that, for any $D \in \mathsf{Defending}(A, \mathcal{E})$ a $B_i \in \{B_1, \ldots, B_n\}$ exists. It therefore follows that $\mathsf{Defending}(A, \mathcal{E}) \subseteq \mathsf{NoDefAgainst}(B_1, \mathcal{E}) \cup \ldots \cup \mathsf{NoDefAgainst}(B_n, \mathcal{E})$.
- Let $A \in \mathcal{E}$ and suppose that $C_1, \ldots, C_j \notin \mathcal{E}$ for some $j \leq k$. By assumption A indirectly attacks C_i for all $i \in \{1, \ldots, j\}$ and since $A \in \mathcal{E}$, C_i is not defended against this attack by A. Therefore $A \in \mathsf{NoDefAgainst}(C_i, \mathcal{E})$. Note that any $D \in \mathsf{Defending}(A, \mathcal{E})$ defends A and therefore indirectly attacks C_i as well. It therefore follows that $D \in \mathsf{NoDefAgainst}(C_i, \mathcal{E})$ and hence $\mathsf{Defending}(A, \mathcal{E}) \subseteq \mathsf{NoDefAgainst}(C_i, \mathcal{E})$ for all $i \in \{1, \ldots, j\}$. □

A.2. Proofs concerning the computation of the explanations

Theorem 1. *Let $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ be an argumentation framework. Then:*

1. *there is an attack-path from A to B of length n iff $n \in \mathsf{Dist}(A, B)$;*
2. *$A \in \mathsf{Reach}(B)$ iff there is an attack-path from A to B;*

In order to show the above theorem, we need some lemmas and propositions. These lemmas and propositions will partly be shown by induction proofs, for which the following remark will be useful.

Remark 15. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A, B \in \mathsf{Args}$. It holds that $A = B$ iff there is an *attack-path* from A to B that has length 0. Similarly, $(A, B) \in \mathsf{Att}$ iff there is an attack-path from A to B of length 1.

Lemma 1. *If $\mathit{ReReach}(A, A', n, \mathsf{S})$ is called, there is an attack-path from A' to A, of length n, along the attacks in* S.

Proof. Suppose that $\mathsf{ReReach}(A, A', n, \mathsf{S})$ is called, either at Line 1 of Algorithm 1 or at Line 2 of Procedure ReReach. We proceed by induction on n.

- **If $n = 0$:** then $\mathsf{ReReach}(A, A, 0, \emptyset)$ is called at Line 1 of Algorithm 1. Since $A = A'$, by Remark 15, there is an attack-path of length 0 from A' to A, without any attacks.
- **If $n = 1$:** then $\mathsf{ReReach}(A, A', n, \mathsf{S})$ was called at the first iteration of Procedure ReReach. Hence $(A', A) \in \mathsf{Att}$ and $(A', A) \notin \emptyset$. By Remark 15, there is an attack-path of length 1 from A' to A.

Suppose now that the claim holds for n up to $k \geq 1$.

- **If $n = k + 1$:** then there is some $B \in \mathsf{Args}$ such that $\mathsf{ReReach}(A, A', n, \mathsf{S})$ is called at Line 2 of the call $\mathsf{ReReach}(A, B, k, \mathsf{S}')$ where $\mathsf{S}' = \mathsf{S} \setminus \{(A', B)\}$. To see that $\mathsf{S}' = \mathsf{S} \setminus \{(A', B)\}$, note that Visited is updated at Line 5 with (A', B) before $\mathsf{ReReach}(A, A', n, \mathsf{S})$ is called and if $(A', B) \in \mathsf{S}'$, the call to $\mathsf{ReReach}(A, A', n, \mathsf{S})$ would not be reached.
  By induction hypothesis, there is an attack-path from B to A of length k along the attacks in S'. Since $(A', B) \notin \mathsf{S}'$, it follows that the attack (A', B) was not used in the attack-path from B to A. Therefore, the path from A' to A along (A', B) and the attack-path B to A is an attack-path from A' to A of length $k + 1$ along the attacks in S. □

The next proposition shows that Algorithm 1 is sound.

Proposition 13. *If $n \in \mathsf{Dist}(A, B)$ then there is an attack-path from A to B of length n.*

Proof. Suppose that $n \in \mathsf{Dist}(A, B)$, that there is an attack-path from A to B of length n is shown by induction on n:

- **If $n = 0$:** then $\mathsf{Dist}(A, B)$ was updated at Line 1 of the algorithm (since at any other place that $\mathsf{Dist}(A, B)$ might be updated, the addition is always more than 0). It follows immediately that $A = B$. Hence there is an attack path from A to B of length 0.
- **If $n = 1$:** then $\mathsf{Dist}(A, B)$ was updated at Line 2 of the procedure in the first iteration of the **for**-loop (since in any other iteration $n \neq 0$). It follows that $(A, B) \in \mathsf{Att}$. Hence the attack-path consists of one attack: (A, B). Thus there is an attack-path from A to B of length 1.

Suppose that the proposition holds for values of n up to k, where $k \geq 1$. Then:

- **If $n = k + 1$:** then $\mathsf{Dist}(A, B)$ was updated at Line 2 of Procedure ReReach. This is only the case if there is some argument $C \in \mathsf{Args}$ such that $\mathrm{ReReach}(B, C, k, \mathsf{S})$ was called and $(A, C) \in \mathsf{Att}$ such that $(A, C) \notin \mathsf{S}$. By induction hypothesis, there is an attack-path from C to B of length k and by Lemma 1, this attack-path is along the attacks in S. Since $(A, C) \notin \mathsf{S}$ by assumption, the path from A to B via the attack (A, C) and the attack-path from C to B is an attack-path, of length $k + 1$.

This shows that for any n, if $n \in \mathsf{Dist}(A, B)$, there is an attack-path of length n from A to B. □

In the next lemma the relation between the attacks in an attack-path and the attacks in Visited in the procedure is shown.

Lemma 2. *If there is an attack-path from A to B, along the attacks $(A, C_1), (C_1, C_2), \ldots, (C_{n-1}, B) \in \mathsf{Att}$, for $C_1, \ldots, C_{n-1} \in \mathsf{Args}$, then during the run of the algorithm $\mathrm{ReReach}(B, A, n, \{(A, C_1), (C_1, C_2), \ldots, (C_{n-1}, B)\})$ will be called.*

Proof. Suppose that there is an attack-path from A to B, along the attacks $(A, C_1), (C_1, C_2), \ldots, (C_{n-1}, B) \in \mathsf{Att}$, for $C_1, \ldots, C_{n-1} \in \mathsf{Args}$. We proceed by induction on $n \geq 1$. Since $B \in \mathsf{Args}$, $\mathrm{ReReach}(B, B, 0, \emptyset)$ will be called at Line 1 of the algorithm.

- **If $n = 1$:** then $(A, B) \in \mathsf{Att}$. At this point Visited is still empty, hence the **for**-loop at Line 2 of Procedure ReReach will be run for A. At Line 2, Visited becomes $\{(A, B)\}$ and at Line 2, $\mathrm{ReReach}(B, A, 1, \{(A, B)\})$ is called.

Suppose the statement holds for values of n up to $k \geq 1$. Then:

- **If $n = k + 1$:** then there are $C_1, \ldots, C_k$ such that there is an attack-path from A to B along the attacks $(A, C_1), \ldots, (C_k, B)$. And hence there is an attack-path from C_1 to B along the attacks $(C_1, C_2), \ldots, (C_k, B)$ of length k. By induction hypothesis, it holds that $\mathrm{ReReach}(B, C_1, k, \{(C_1, C_2), \ldots, (C_k, B)\})$ is called during the run of the algorithm.
  Since $(A, C_1) \in \mathsf{Att}$, A is one of the arguments considered during this call. By assumption $(A, C_1), \ldots, (C_k, B)$ is an attack-path from A to B, therefore no attack appears twice. Hence $(A, C_1) \notin \{(C_1, C_2), \ldots, (C_k, B)\}$. Then Visited will be updated with (A, C_1) at Line 2 and at Line 2 $\mathrm{ReReach}(B, A, k + 1, \{(A, C_1), (C_1, C_2), \ldots, (C_k, B)\})$ is called.

This shows that for any n, if there is an attack-path from A to B along the attacks in S, $\mathrm{ReReach}(B, A, n, \mathsf{S})$ will be called. □

The next proposition shows that Algorithm 1 is complete.

Proposition 14. *If there is an attack-path from A to B of length n then $n \in \mathsf{Dist}(A, B)$.*

Proof. Suppose that there is an attack-path from A to B of length n. We proceed again by induction on n.

- **If $n = 0$:** then by Remark 15 $A = B$. By Line 1 of Algorithm 1, $0 \in \mathsf{Dist}(A, A)$ and $0 \in \mathsf{Dist}(B, B)$.
- **If $n = 1$:** then by Remark 15 $(A, B) \in \mathsf{Att}$. By Line 2 of Procedure ReReach, $1 \in \mathsf{Dist}(A, B)$.

Suppose that the statement holds for n up to $k \geq 1$.

- **If $n = k + 1$:** then there are $C_1, \ldots, C_{k+1} \in \mathsf{Args}$, such that $A = C_1$, $B = C_{k+1}$, $(C_1, C_2), \ldots, (C_k, C_{k+1}) \in \mathsf{Att}$ and there are no $1 \leq i, j \leq k$ such that $i \neq j$ and $(C_i, C_{i+1}) = (C_j, C_{j+1})$ (i.e., the attack-path does not follow an attack twice). Note that for any $2 \leq i, j \leq k + 1$ such that $i \leq j$, the corresponding subset of attacks is an attack-path from C_i to C_j. In particular, $(C_2, C_3), \ldots, (C_k, B)$ is an attack-path from C_2 to B of length k.
  By induction hypothesis, $k \in \mathsf{Dist}(C_2, B)$. Since $k \geq 1$, $\mathsf{Dist}(C_2, B)$ was updated at Line 2 of Procedure ReReach during the $\mathrm{ReReach}(B, C_3, k - 1, \mathsf{S}')$ call of the procedure. By Lemma 2 it follows that $\mathsf{S}' = \{(C_3, C_4), \ldots, (C_k, B)\}$. Then, at Line 2 Visited

is updated with (C_2, C_3) and, at Line 2, ReReach$(B, C_2, k, \mathsf{S}' \cup \{(C_2, C_3)\})$ is called. Since $(C_1, C_2) \in \mathsf{Att}$ and since there is an attack-path from C_1 to C_{k+1} along the attacks of $\mathsf{S}' \cup \{(C_1, C_2), (C_2, C_3)\}$, Dist$(A, B)$ will be updated at Line 2 with $k+1$. □

In our paper we are interested in the distance between two arguments (since this determines whether the relation is an attack (the distance is odd) or a defense (the distance is even)), but also in the arguments from which an argument is reachable. Both are computed by Algorithm 1 and the next lemma shows the relation between the two.

Lemma 3. Dist$(A, B) \neq \emptyset$ *iff* $A \in$ Reach(B).

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A, B \in \mathsf{Args}$. Assume that Algorithm 1 was run on $\mathcal{AF}$. Consider both directions separately.

⇒ Suppose that Dist$(A, B) \neq \emptyset$. This direction is shown by induction on the minimal value n in Dist(A, B).
- **If** $n = 0$**:** then Dist(A, B) was updated at Line 1 of the algorithm (since at Line 2 of the procedure the addition is always more than 0) and thus $A = B$. By Line 1 again it follows that $A \in$ Reach(A).
- **If** $n = 1$**:** then Dist(A, B) was updated at Line 2 of the procedure during the first iteration of the **for**-loop, in which case $(A, B) \in \mathsf{Att}$. By Line 2 it follows that $A \in$ Reach(B).

Now suppose that the statement holds for n up to a value of $k \geq 1$.
- **If** $n = k + 1$**:** then Dist(A, B) was updated at Line 2 of Procedure ReReach. Therefore, during this run of ReReach, at Line 2, Reach(B) is updated with Reach(A). Note that by Line 1 of Algorithm 1 $A \in$ Reach(A) and hence $A \in$ Reach(B).

⇐ Now assume that $A \in$ Reach(B). We consider three cases:
- $A = B$, then Reach(B) was updated at Line 1 of Algorithm 1, such that $B \in$ Reach(B) and $0 \in$ Dist(A, B).
- Reach(B) was updated at Line 2 of Procedure ReReach, with Reach(A). Then at Line 2, Dist(A, B) is updated with $n + 1$.
- Reach(B) was updated at Line 2 of Procedure ReReach, with Reach(C) and $A \in$ Reach(C). Hence, by Proposition 13, there is an attack-path from A to C and there is an attack-path from C to B. If no attack in the path from C to B is used in the attack-path from A to C, the procedure will call all arguments in the attack-path from A to C until it reaches A. At which point Dist(A, B) will be updated.
  Suppose now that there is some $(D_1, D_2) \in \mathsf{Att}$, such that (D_1, D_2) appears in both paths. Then there is an attack-path from A to D_1 (along the attacks $(A, E_k), \ldots, (E_1, D_1)$, where $E_k, \ldots, E_1 \in \mathsf{Args}$) and there is an attack-path from D_1 to B. Without loss of generality, suppose that (D_1, D_2) is such that there is no attack (D_1', D_2') in the attack-path from A to D_1 that also appears in the attack-path from C to B (otherwise the described procedure has to be repeated). Since there is an attack-path from D_1 to B (say of length l_d), by Lemma 2, ReReach$(B, D_1, l_d, \mathsf{S})$ is called. By assumption $\{(A, E_k), \ldots, (E_1, D_1)\} \cap \mathsf{S} = \emptyset$. Hence, for each $i \in \{1, \ldots, k\}$, during the call for ReReach$(B, D_1, l_d, \mathsf{S})$, ReReach$(B, E_i, l_d + i, \mathsf{S} \cup \{(E_i, E_{i-1}), \ldots, (E_1, D_1)\})$ is called. At ReReach$(B, E_k, l_d + k, \mathsf{S} \cup \{(E_k, E_{k-1}), \ldots, (E_1, D_1)\})$, note that $(A, E_k) \notin \mathsf{S} \cup \{(E_k, E_{k-1}), \ldots, (E_1, D_1)\}$. Hence Reach$(B)$ is updated with Reach(A) and Dist(A, B) is updated with $l_d + k + 1$. Therefore Dist$(A, B) \neq \emptyset$.

This shows that, in any situation, if $A \in$ Reach(B) then Dist$(A, B) \neq \emptyset$. □

With the above results we have the proof of Theorem 1:

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework, $A, B \in \mathsf{Args}$ and suppose that Algorithm 1 was run on $\mathcal{AF}$. Then:

1. Soundness and completeness of the algorithm follows immediately by Propositions 13 and 14.
2. By Lemma 3 we know that $A \in$ Reach(B) iff Dist$(A, B) \neq \emptyset$ and by the soundness and completeness of the algorithm (i.e., the first item) it is known that $n \in$ Dist(A, B) iff there is an atack-path from A to B. □

We now turn to the computational complexity of the algorithm. Note that the algorithm does not determine whether an argument is accepted or not. It is therefore important that the extensions have been determined before running the algorithm.

Theorem 2. *Algorithm 1 runs in polynomial time. In particular the time complexity is* $\mathcal{O}(|\mathsf{Args}| \cdot |\mathsf{Att}|^2)$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and suppose that Algorithm 1 was run on this framework. Then:

- The first **for**-call of the algorithm takes |Args| time.
- Procedure ReReach runs in $|\mathsf{Att}|^2$: from each attack at most all other attacks are visited exactly once (|Att|) and at most |Att| attacks end in a single argument (|Att|).
- The procedure is called |Args| times from Algorithm 1.

This gives a total of $|\mathsf{Args}| + |\mathsf{Args}| \cdot |\mathsf{Att}|^2$, assuming that Att $\neq \emptyset$ (this is safe to assume since argumentation could be considered interesting only when there are attacks), $\mathcal{O}(|\mathsf{Args}| \cdot |\mathsf{Att}|^2)$. □

A.3. Proofs of the results on the variations of $\mathbb{D}$

Corollary 1. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an argumentation framework, $A \in Args$ and let $\mathbb{D} = \mathsf{DirDefending}$. Then:*

1. $\mathsf{GrdAcc}(A) \subseteq \bigcap \mathsf{SemAcc}^{\cap}(A)$ *for* $\mathsf{Sem} \in \{\mathsf{Grd}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$.
2. $\mathsf{StbAcc}^{\cap}(A) \subseteq \mathsf{SstbAcc}^{\cap}(A) \subseteq \mathsf{PrfAcc}^{\cap}(A) \subseteq \mathsf{CmpAcc}^{\cap}(A)$.
3. *For each* $\mathsf{S} \in \mathsf{CmpAcc}^{\cap}(A)$ *there is an* $\mathsf{S}' \in \mathsf{PrfAcc}^{\cap}(A)$ *such that* $\mathsf{S} \subseteq \mathsf{S}'$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework and let $A \in \mathsf{Args}$. Suppose that A is accepted under Sem and $\star \in \{\cap, \cup\}$. The proof is similar to the proof of Proposition 1. This follows since any direct defender B of A in a Sem_1-extension $\mathcal{E}$ is also a direct defender of A in any Sem_2-extension that contains $\mathcal{E}$. □

Corollary 2. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an AF, let $A \in Args$ be such that A is accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ *and* $\star \in \{\cap, \cup\}$ *and let $\mathbb{D} = \mathsf{DirDefending}$. Then* $\mathsf{SemAcc}^{\cap}(A) = \{\emptyset\}$ *and* $\mathsf{SemAcc}^{\cup}(A, \mathcal{E}) = \emptyset$ *for any* $\mathcal{E} \in \mathsf{SemWith}(A)$ *iff there is no* $B \in Args$ *such that* $(B, A) \in Att$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an AF, let $A \in \mathsf{Args}$ be such that A is accepted w.r.t. $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ and $\star \in \{\cap, \cup\}$ and let $\mathbb{D} = \mathsf{DirDefending}$.

- ⇒ Suppose that $\mathsf{SemAcc}^{\cap} = \{\emptyset\}$ or $\mathsf{SemAcc}^{\cup}(A, \mathcal{E}) = \emptyset$ for any $\mathcal{E} \in \mathsf{SemWith}(A)$. The proof follows similarly to the proof for $\mathbb{D} = \mathsf{Defending}$ in Proposition 3.
- ⇐ If A is not attacked, it follows that $\mathsf{SemAcc}^{\cap}(A, \mathcal{E}) = \{\emptyset\}$ for any $\mathcal{E} \in \mathsf{SemWith}(A)$ and $\mathsf{SemAcc}^{\cup}(A) = \emptyset$. □

Corollary 3. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an argumentation framework, $A \in Args$ be non-accepted, $\star \in \{\cap, \cup\}$ and let $\mathbb{D} \in \{$*NoDirDefense, NoSelfDefense*$\}$. Then:*

- $\mathsf{GrdNotAcc}(A) \subseteq \bigcap \mathsf{CmpNotAcc}^{\cup}(A)$.
- $\mathsf{SstbNotAcc}^{\cup}(A) \subseteq \mathsf{PrfNotAcc}^{\cup}(A) \subseteq \mathsf{CmpNotAcc}^{\cup}(A)$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework, $A \in \mathsf{Args}$ be non-accepted and let $\star \in \{\cap, \cup\}$. Recall that the proof for Proposition 4 is based on the relation between the extensions of $\mathcal{AF}$ under the different semantics. This relation is not changed by the variations of $\mathbb{D}$. The proof remains therefore the same. □

Corollary 4. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an argumentation framework and let $A \in Args$ be such that A is non-accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ *and* $\star \in \{\cap, \cup\}$. *Then:*

- *if* $\mathbb{D} = \mathsf{NoDirDefense}$ *then* $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) \neq \emptyset$ *for any extension* $\mathcal{E} \in \mathsf{SemWithout}(A)$ *and* $\mathsf{SemNotAcc}^{\cup}(A) \neq \{\emptyset\}$*; and*
- *if* $\mathbb{D} = \mathsf{NoSelfDefense}$ *then* $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) = \emptyset$ *for any extension* $\mathcal{E} \in \mathsf{SemWithout}(A)$ *or* $\mathsf{SemNotAcc}^{\cup}(A) = \{\emptyset\}$ *implies that for all* $B \in Args$ *such that* $(B, A) \in Att$, $(A, B) \in Att$ *as well.*

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an argumentation framework and let $A \in \mathsf{Args}$ be such that A is non-accepted w.r.t. $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$ and $\star \in \{\cap, \cup\}$. Because of our choice of semantics (i.e., completeness-based) and the fact that A is not accepted, there is some $B \in \mathsf{Args}$, such that $(B, A) \in \mathsf{Att}$ and A is not defended by some $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$. Consider both items:

- let $\mathbb{D} = \mathsf{NoDirDefense}$. It follows immediately that $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) \neq \emptyset$ for any $\mathcal{E} \in \mathsf{SemWithout}(A)$ and $\mathsf{SemNotAcc}^{\cup}(A) \neq \{\emptyset\}$.
- let $\mathbb{D} = \mathsf{NoSelfDefense}$. Suppose that $\mathsf{SemNotAcc}^{\cap}(A, \mathcal{E}) = \emptyset$ for any $\mathcal{E} \in \mathsf{SemWithout}(A)$ [respectively $\mathsf{SemNotAcc}^{\cup}(A) = \{\emptyset\}$]. Then, by Proposition 5, it follows that $\mathsf{NoDefAgainst}(A, \mathcal{E}) \setminus \mathsf{NoSelfDefense}(A, \mathcal{E}) = \emptyset$ [for all extensions $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$]. It follows that $(A, B) \in \mathsf{Att}$, for each attacker of A. □

A.4. Proofs concerning necessity and sufficiency of explanations

Proposition 8. *Let $\mathcal{AF} = \langle Args, Att\rangle$ be an argumentation framework and let $A \in Args$ be accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ *and* $\star \in \{\cap, \cup\}$. *Then:*

- *for all* $\mathcal{E} \in \mathsf{SemWith}(A)$, $\mathsf{Defending}(A, \mathcal{E}) \in \mathsf{Suff}(A, \emptyset)$*;*
- $\bigcap_{\mathcal{E} \in \mathsf{SemWith}(A)} \mathsf{Defending}(A, \mathcal{E}) = \mathsf{Nec}(A, \emptyset)$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att}\rangle$ be an AF and let $A \in \mathsf{Args}$ be an argument that is accepted w.r.t. Sem. Consider both items.

- Since A is accepted, there is some $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ such that $A \in \mathcal{E}$. Let $\mathsf{T} = \mathsf{Defending}(A, \mathcal{E})$. By definition, T is relevant for A (i.e., all $B \in \mathsf{T}$ (in)directly defend A and since $B \in \mathcal{E}$, $(B, B) \notin \mathsf{Att}$). Now suppose that there is some $C \in \mathsf{Args}$ such that C attacks A and A is not defended by T. By assumption $A \in \mathcal{E}$. Hence there is a $D \in \mathcal{E}$ such that $(D, C) \in \mathsf{Att}$. But then D (in)directly defends A and therefore $D \in \mathsf{T}$. Thus T defends A against all its attackers and therefore $\mathsf{T} \in \mathsf{Suff}(A, \emptyset)$.
- Let $\mathsf{T} = \bigcap_{\mathcal{E} \in \mathsf{SemWith}(A)} \mathsf{Defending}(A, \mathcal{E})$, since A is accepted it follows that $\mathsf{SemWith}(A) \neq \emptyset$. Suppose there is some $B \in \mathsf{T}$ which is not necessary for the acceptance of A. Then there is an $\mathcal{E} \in \mathsf{SemWith}(A)$ such that $B \notin \mathcal{E}$. However, by definition of T, $B \in \bigcap \mathsf{SemWith}(A)$. Hence T is necessary for A. To see that T contains all the necessary arguments, assume it does not. Then there is some $B \in \mathsf{Args}$ such that $B \notin \mathsf{T}$ but B is necessary for the acceptance of A. However, since $B \notin \mathsf{T}$, there is some $\mathcal{E} \in \mathsf{SemWith}(A)$ such that $B \notin \mathcal{E}$, but $A \in \mathcal{E}$. A contradiction. □

Proposition 7. *Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and let $A \in \mathsf{Args}$ be accepted w.r.t. some $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ and $\star \in \{\cap, \cup\}$. Then:*

1. *For all $\mathsf{S} \in \mathsf{Suff}(A, \emptyset)$, $\mathsf{S} \in \mathsf{Adm}(\mathcal{AF})$;*
2. *$\mathsf{Suff}(A, \emptyset) = \{A\}$ iff there is no $B \in \mathsf{Args}$ such that $(B, A) \in \mathsf{Att}$ or $\{A\}$ defends itself against all its attackers;*
3. *$\mathsf{Nec}(A, \emptyset) = \{A\}$ iff there is no $B \in \mathsf{Args}$ such that $(B, A) \in \mathsf{Att}$, or $\{A\}$ defends itself against all its attackers or it holds that $\bigcap \mathsf{Suff}(A, \emptyset) = \{A\}$.*
4. *$\mathsf{Nec}(A, \emptyset) = \bigcap \mathsf{Suff}(A, \emptyset)$.*

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A \in \mathsf{Args}$ be accepted w.r.t. some Sem.

1. Let $\mathsf{S} \in \mathsf{Suff}(A)$. By definition S is conflict-free and defends S against all its attackers. It follows that $\mathsf{S} \in \mathsf{Adm}(\mathcal{AF})$.
2. Suppose that $\mathsf{Suff}(A, \emptyset) = \{A\}$. Then $\{A\}$ is sufficient for the acceptance of A, from which it follows that A attacks all its attackers, or $\emptyset$ is sufficient for the acceptance of A. In the latter case, there is no $\mathsf{S} \subseteq \mathsf{Args}$ such that S is relevant for A and defends A against all its arguments. Since A is accepted by assumption, it follows that A is not attacked at all. Now suppose that there is no $B \in \mathsf{Args}$ such that $(B, A) \in \mathsf{Att}$. Then there is no $\mathsf{S} \subseteq \mathsf{Args}$ that is relevant for A and hence $\emptyset$ is sufficient for the acceptance of A and therefore $\mathsf{Suff}(A, \emptyset) = \{A\}$.
3. First suppose that $\mathsf{Nec}(A, \emptyset) = \{A\}$. Then there is no argument relevant for A, other than possibly A itself (from which it follows that there is no $B \in \mathsf{Args}$ such that $(B, A) \in \mathsf{Att}$ or if $(B, A) \in \mathsf{Att}$ then $(A, B) \in \mathsf{Att}$) or there is no $B \in \mathsf{Args} \setminus \{A\}$ such that $B \in \bigcap \mathsf{SemWith}(A)$. Note that for each $\mathsf{S} \in \mathsf{Suff}(A, \emptyset)$ there is some $\mathcal{E} \in \mathsf{SemWith}(A)$ such that $\mathsf{S} \subseteq \mathcal{E}$. Since $\bigcap \mathsf{SemWith}(A) = \{A\}$ it follows that $\bigcap \mathsf{Suff}(A, \emptyset) = \{A\}$ as well.
   For the other direction suppose first that A is not attacked at all, then there is no argument relevant for A from which it follows that $\mathsf{Nec}(A, \emptyset) = \{A\}$. Now suppose that $\mathsf{Suff}(A, \emptyset) = \{A\}$. By assumption A is accepted and A is attacked, hence $\mathsf{Suff}(A, \emptyset) = \{A\}$ iff A attacks all its attackers. It follows that for each $\mathsf{S} \in \mathsf{Suff}(A, \emptyset)$ and for each $B \in \mathsf{S}$ there is an $\mathsf{S}' \in \mathsf{Suff}(A, \emptyset)$ such that $B \notin \mathsf{S}'$ and therefore also an $\mathcal{E} \in \mathsf{Adm}(\mathcal{AF})$ with $B \notin \mathcal{E}$ but $A \in \mathcal{E}$. Therefore none of the arguments is necessary: $\mathsf{Nec}(A, \emptyset) = \{A\}$.
4. In view of the above two items, suppose that A is attacked by some argument. Let $B \in \mathsf{Nec}(A, \emptyset)$ and suppose that $B \notin \bigcap \mathsf{Suff}(A, \emptyset)$. Then there is some $\mathsf{S} \in \mathsf{Suff}(A, \emptyset)$ such that $B \notin \mathsf{S}$. Note that $\mathsf{S} \in \mathsf{Adm}(\mathcal{AF})$. However, $B \notin \mathsf{S}$, a contradiction with $B \in \mathsf{Nec}(A, \emptyset)$. □

Proposition 9. *Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an argumentation framework and let $A \in \mathsf{Args}$ be non-accepted w.r.t. $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ and $\star \in \{\cap, \cup\}$. Then:*

- *$\mathsf{SuffNot}(A, \emptyset) \neq \emptyset$;*
- *$\mathsf{NecNot}(A, \emptyset) = \emptyset$ implies that there are at least two direct attackers of A.*

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A \in \mathsf{Args}$ be an argument that is not accepted w.r.t. $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Sstb}\}$.

- Suppose that $\mathsf{SuffNot}(A) = \emptyset$. Then there is no $\mathsf{S} \subseteq \mathsf{Args}$ that is relevant for A and in which $B \in \mathsf{Args}$ (in)directly attacks A. It follows that there is no $B \in \mathsf{Args}$ such that $(B, A) \in \mathsf{Att}$. A contradiction with the assumption that A is non-accepted and that $(A, A) \notin \mathsf{Att}$.
- It follows that there are $B_1, \ldots, B_n \in \mathsf{Args}$ such that $(B_1, A), \ldots, (B_n, A) \in \mathsf{Att}$. Assume that $\mathsf{NecNot}(A, \emptyset) = \emptyset$ but that $n = 1$. Since by assumption in this section $(A, A) \notin \mathsf{Att}$, it follows that A is not attacked in $\mathcal{AF}_{\downarrow B_1}$ and should therefore be accepted in any complete extension. Hence $n \geq 2$. □

Lemma 4. *Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$, $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for some $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ and $A \in \mathsf{Args}$. If there is a $B \in \mathcal{E}$ such that $(B, A) \in \mathsf{Att}$, then $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF}_{\downarrow A})$.*[20]

[20] For $\mathsf{Sem} = \mathsf{Grd}$ this lemma was shown in [14].

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$, $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ for some Sem and $A, B \in \mathsf{Args}$ such that $B \in \mathcal{E}$ and $(B, A) \in \mathsf{Att}$. Note that $\mathcal{E}$ is still admissible in $\mathcal{AF}_{\downarrow A}$ since no new attacks are added.

$\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Prf}\}$. Now suppose there is some $C \in \mathsf{Args}$ such that $C \notin \mathcal{E}$ but C is defended by $\mathcal{E}$ in $\mathcal{AF}_{\downarrow A}$. If C is not attacked at all in $\mathcal{AF}_{\downarrow A}$, since $C \notin \mathcal{E}$, $(A, C) \in \mathsf{Att}$, but then $\mathcal{E}$ defends C in $\mathcal{AF}$, a contradiction. Hence there is some $D \in \mathsf{Args}$ such that $(D, C) \in \mathsf{Att}$ and $\mathcal{E}$ defends against this attack in $\mathcal{AF}_{\downarrow A}$, but then $\mathcal{E}$ would defend C in $\mathcal{AF}$ as well. Again a contradiction. Hence $\mathcal{E}$ is complete in $\mathcal{AF}_{\downarrow A}$ and if $\mathcal{E}$ was maximally complete in $\mathcal{AF}$ it is still maximally complete in $\mathcal{AF}_{\downarrow A}$.

$\mathsf{Sem} = \mathsf{Sstb}$. Any argument, other than A, attacked by $\mathcal{E}$ is still attacked by $\mathcal{E}$ in $\mathcal{AF}_{\downarrow A}$. Since $\mathcal{E}$ is still complete, it follows that $\mathcal{E}$ is also still semi-stable. □

Proposition 10. *Let $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ be an argumentation framework and let $A \in \mathit{Args}$ be an argument that is not accepted w.r.t.* $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Stb}\}$ *and* $\star \in \{\cap, \cup\}$. *Then:*

- *for all* $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF})$ *such that* $A \notin \mathcal{E}$, $\mathsf{NoDefAgainst}(A, \mathcal{E}) \in \mathsf{SuffNot}(A, \emptyset)$;
- $\mathsf{NecNot}(A, \emptyset) \subseteq \bigcap_{\mathcal{E} \in \mathsf{SemWithout}(A)} \mathsf{NoDefAgainst}(A, \mathcal{E})$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A \in \mathsf{Args}$ be non-accepted w.r.t. $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Sstb}\}$. Consider both items:

- By definition of NoDefAgainst, $\mathsf{T} = \mathsf{NoDefAgainst}(A, \mathcal{E})$ is relevant for A. We show that there is a $B \in \mathsf{T}$ such that $(B, A) \in \mathsf{Att}$. Suppose there is no such B, then A is not attacked at all or $\mathcal{E}$ defends A against all its direct attackers and therefore against all its attackers, both are a contradiction with the completeness of $\mathcal{E}$. Hence there is such a $B \in \mathsf{T}$. From which it follows that $\mathsf{NoDefAgainst}(A, \mathcal{E}) \in \mathsf{SuffNot}(A)$.
- Let $B \in \mathsf{NecNot}(A)$ and suppose that $B \notin \bigcap_{\mathcal{E} \in \mathsf{SemWithout}(A)} \mathsf{NoDefAgainst}(A)$. Then there is some $\mathcal{E} \in \mathsf{SemWithout}(A)$ such that $B \notin \mathsf{NoDefAgainst}(A, \mathcal{E})$. By assumption, B is relevant for A and thus (in)directly attacks A. From which it follows that there is some $C \in \mathcal{E}$ such that $(C, B) \in \mathsf{Att}$. By Lemma 4, $\mathcal{E} \in \mathsf{Sem}(\mathcal{AF}_{\downarrow B})$, a contradiction with the assumption that $B \in \mathsf{NecNot}(A)$. □

Proposition 11. *Let $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ be an argumentation framework, let $A \in \mathit{Args}$ and let* $\preceq \in \{\subseteq, \leq\}$. *Then, for* $\mathbb{D} = \mathsf{Defending}$ *and* $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Stb}, \mathsf{Sstb}, \mathsf{Egr}\}$:

- *for every* $\mathsf{S} \in \mathsf{SemAcc}^{\cap}(A)$ *there is an* $\mathsf{S}' \in \mathsf{SemAcc}^{\cap}_{\preceq}(A)$ *for* $\mathbb{D} = \mathsf{MinSuff}^{\preceq}$ *such that* $\mathsf{S}' \preceq \mathsf{S}$;
- *where* $\mathsf{Sem} = \mathsf{Adm}$, *for every* $\mathsf{S} \in \mathsf{SemAcc}^{\cap}_{\preceq}(A)$ *where* $\mathbb{D} = \mathsf{MinSuff}^{\preceq}$ *also* $\mathsf{S} \in \mathsf{SemAcc}^{\cap}(A)$;
- *for all* $\mathsf{S} \in \mathsf{SemAcc}^{\cap}(A)$, $\mathsf{Nec}(A, \emptyset) \subseteq \mathsf{S}$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A \in \mathsf{Args}$ be accepted w.r.t. $\mathsf{Sem} \in \{\mathsf{Adm}, \mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Sstb}\}$:

- Let $\mathsf{S} \in \mathsf{SemAcc}^{\cap}(A)$ and $\mathbb{D} = \mathsf{Defending}$, then $\mathsf{S} = \mathsf{Defending}(A, \mathcal{E})$ for some $\mathcal{E} \in \mathsf{SemWith}(A)$. By Proposition 8 it follows that $\mathsf{S} \in \mathsf{Suff}(A, \emptyset)$. Then there is some $\mathsf{S}' \in \mathsf{MinSuff}^{\preceq}(A, \emptyset)$ such that $\mathsf{S}' \preceq \mathsf{S}$ and $\mathsf{S}' \in \mathsf{SemAcc}^{\cap}_{\preceq}(A)$ for $\mathbb{D} = \mathsf{MinSuff}^{\preceq}$, for any of the considered semantics.
- Let $\mathsf{Sem} = \mathsf{Adm}$ and $\mathsf{S} \in \mathsf{MinSuff}^{\preceq}(A, \emptyset)$. By Proposition 7, $\mathsf{S} \in \mathsf{Adm}(\mathcal{AF})$ and by definition of a sufficient set of arguments, S defends A against all its attackers. Therefore $\mathsf{Defending}(A, \mathsf{S}) = \mathsf{S}$. Suppose now that S is such that $\mathsf{Defending}(A, \mathsf{S})$ is not $\preceq$-minimal among $\mathcal{E} \in \mathsf{Adm}(\mathcal{AF})$. Then there is some $\mathcal{E} \in \mathsf{AdmWith}(A)$ such that $\mathsf{Defending}(A, \mathcal{E}) \prec \mathsf{Defending}(A, \mathsf{S})$. By Proposition 8, $\mathsf{Defending}(A, \mathcal{E}) \in \mathsf{Suff}(A, \emptyset)$. A contradiction since $\mathsf{S} \in \mathsf{MinSuff}^{\preceq}(A, \emptyset)$ and $\mathsf{Defending}(A, \mathcal{E}) \prec \mathsf{S}$.
- This follows immediately from the second item in Proposition 8. □

Proposition 12. *Let $\mathcal{AF} = \langle \mathit{Args}, \mathit{Att} \rangle$ be an argumentation framework, let $A \in \mathit{Args}$ and let* $\preceq \in \{\leq, \subseteq\}$. *Then, for* $\mathbb{D} = \mathsf{NoDefAgainst}$ *and* $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Prf}, \mathsf{Idl}, \mathsf{Sstb}, \mathsf{Egr}\}$:

- *for every* $\mathsf{S} \in \mathsf{SemNotAcc}^{\cup}(A)$ *there is an* $\mathsf{S}' \in \mathsf{SuffNotAcc}^{\cup}_{\preceq}(A)$ *where* $\mathbb{D} = \mathsf{MinSuffNot}^{\preceq}$ *such that* $\mathsf{S}' \preceq \mathsf{S}$.
- *for all* $\mathsf{S} \in \mathsf{SemNotAcc}^{\cup}(A)$, $\mathsf{NecNot}(A, \emptyset) \subseteq \mathsf{S}$.

Proof. Let $\mathcal{AF} = \langle \mathsf{Args}, \mathsf{Att} \rangle$ be an AF and let $A \in \mathsf{Args}$ be non-accepted w.r.t. $\mathsf{Sem} \in \{\mathsf{Cmp}, \mathsf{Grd}, \mathsf{Prf}, \mathsf{Sstb}\}$:

- Let $\mathsf{S} \in \mathsf{SemNotAcc}^{\cup}(A)$ for $\mathbb{D} = \mathsf{NoDefAgainst}$. Note that we have $\mathsf{S} = \mathsf{NoDefAgainst}(A, \mathcal{E})$ for some $\mathcal{E} \in \mathsf{SemWithout}(A)$. Hence, by Proposition 10 it follows that $\mathsf{S} \in \mathsf{SuffNot}(A, \emptyset)$. Therefore, there is some $\mathsf{S}' \in \mathsf{MinSuffNot}^{\preceq}(A, \emptyset)$ such that $\mathsf{S}' \preceq \mathsf{S}$, for any of the considered semantics.
- This follows from the second item in Proposition 10. □

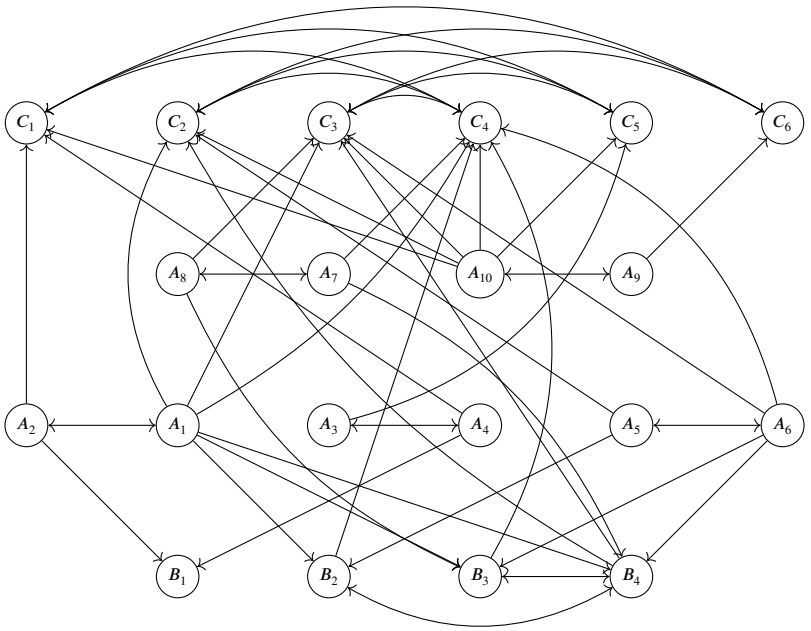


Fig. 10. Graphical representation of the framework $\mathcal{AF}_9$ on online trade fraud.

Appendix B. Formal construction explanations online trade fraud

In this appendix we collect the preferred extensions and provide a formal construction and an exhaustive list of the acceptance explanations of the arguments C_1, C_2, C_3, C_4, C_5 and C_6. See Fig. 10 for a graphical representation of the argumentation framework. The argumentation framework $\mathcal{AF}_9$ has 35 preferred extensions:

$$\mathcal{E}_1 = \{A_1, A_3, A_5, A_7, A_9, B_1, C_1\} \qquad \mathcal{E}_2 = \{A_1, A_3, A_5, A_7, A_{10}, B_1, C_6\}$$
$$\mathcal{E}_3 = \{A_1, A_3, A_5, A_8, A_9, B_1, C_1\} \qquad \mathcal{E}_4 = \{A_1, A_3, A_5, A_8, A_{10}, B_1, C_6\}$$
$$\mathcal{E}_5 = \{A_1, A_3, A_6, A_7, A_9, B_1, C_1\} \qquad \mathcal{E}_6 = \{A_1, A_3, A_6, A_7, A_{10}, B_1, C_6\}$$
$$\mathcal{E}_7 = \{A_1, A_3, A_6, A_8, A_9, B_1, C_1\} \qquad \mathcal{E}_8 = \{A_1, A_3, A_6, A_8, A_{10}, B_1, C_6\}$$
$$\mathcal{E}_9 = \{A_1, A_4, A_5, A_7, A_9, C_5\} \qquad \mathcal{E}_{10} = \{A_1, A_4, A_5, A_7, A_{10}, C_6\}$$
$$\mathcal{E}_{11} = \{A_1, A_4, A_5, A_8, A_9, C_5\} \qquad \mathcal{E}_{12} = \{A_1, A_4, A_5, A_8, A_{10}, C_6\}$$
$$\mathcal{E}_{13} = \{A_1, A_4, A_6, A_7, A_9, C_5\} \qquad \mathcal{E}_{14} = \{A_1, A_4, A_6, A_7, A_{10}, C_6\}$$
$$\mathcal{E}_{15} = \{A_1, A_4, A_6, A_8, A_9, C_5\} \qquad \mathcal{E}_{16} = \{A_1, A_4, A_6, A_8, A_{10}, C_6\}$$
$$\mathcal{E}_{17} = \{A_2, A_3, A_5, A_7, A_9, B_3, C_3\} \qquad \mathcal{E}_{18} = \{A_2, A_3, A_5, A_7, A_{10}, B_3, C_6\}$$
$$\mathcal{E}_{19} = \{A_2, A_3, A_5, A_8, A_9, B_4, C_4\} \qquad \mathcal{E}_{20} = \{A_2, A_3, A_5, A_8, A_{10}, B_4, C_6\}$$
$$\mathcal{E}_{21} = \{A_2, A_3, A_6, A_7, A_9, B_2, C_2\} \qquad \mathcal{E}_{22} = \{A_2, A_3, A_6, A_7, A_{10}, B_2, C_6\}$$
$$\mathcal{E}_{23} = \{A_2, A_3, A_6, A_8, A_9, B_2, C_2\} \qquad \mathcal{E}_{24} = \{A_2, A_3, A_6, A_8, A_{10}, B_2, C_6\}$$
$$\mathcal{E}_{25} = \{A_2, A_4, A_5, A_7, A_9, B_3, C_3\} \qquad \mathcal{E}_{26} = \{A_2, A_4, A_5, A_7, A_9, B_3, C_5\}$$
$$\mathcal{E}_{27} = \{A_2, A_4, A_5, A_7, A_{10}, B_3, C_6\} \qquad \mathcal{E}_{28} = \{A_2, A_4, A_5, A_8, A_9, B_4, C_4, C_5\}$$
$$\mathcal{E}_{29} = \{A_2, A_4, A_5, A_8, A_{10}, B_4, C_6\} \qquad \mathcal{E}_{30} = \{A_2, A_4, A_6, A_7, A_9, B_2, C_2\}$$
$$\mathcal{E}_{31} = \{A_2, A_4, A_6, A_7, A_9, B_2, C_5\} \qquad \mathcal{E}_{32} = \{A_2, A_4, A_6, A_7, A_{10}, B_2, C_6\}$$
$$\mathcal{E}_{33} = \{A_2, A_4, A_6, A_8, A_9, B_2, C_2\} \qquad \mathcal{E}_{34} = \{A_2, A_4, A_6, A_8, A_9, B_2, C_5\}$$
$$\mathcal{E}_{35} = \{A_2, A_4, A_6, A_8, A_{10}, B_2, C_6\}.$$

In the discussion of the example we are interested in explanations for *it is (not) a case of fraud*: the (non-)acceptance of the arguments C_1, C_2, C_3, C_4, C_5 and C_6. We first provide the sets Defending for each of these arguments:

$$\mathsf{Defending}(C_1) = \{A_1, A_2, A_3, A_5, A_6, A_7, A_9, A_{10}, B_2, B_3, C_1, C_2, C_3\}$$
$$\mathsf{Defending}(C_2) = \{A_1, A_2, A_3, A_5, A_6, A_7, A_9, A_{10}, B_2, B_3, C_1, C_2, C_3\}$$
$$\mathsf{Defending}(C_3) = \{A_1, A_2, A_3, A_5, A_6, A_7, A_9, A_{10}, B_2, B_3, C_1, C_2, C_3\}$$
$$\mathsf{Defending}(C_4) = \{A_1, A_2, A_4, A_5, A_6, A_8, A_9, A_{10}, B_4, C_4, C_5, C_6\}$$
$$\mathsf{Defending}(C_5) = \{A_1, A_2, A_4, A_5, A_6, A_8, A_9, A_{10}, B_4, C_4, C_5, C_6\}$$
$$\mathsf{Defending}(C_6) = \{A_1, A_2, A_4, A_5, A_6, A_8, A_9, A_{10}, B_4, C_4, C_5, C_6\}.$$

Based on the above sets we can now determine the basic argument explanations for their credulous acceptance:

$$\mathsf{Prf\,Acc}^{\cup}(C_1, \mathcal{E}_1) = \{A_1, A_3, A_5, A_7, A_9, C_1\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_1, \mathcal{E}_3) = \{A_1, A_3, A_5, A_9, C_1\}$$
$$\mathsf{Prf\,Acc}^{\cup}(C_1, \mathcal{E}_5) = \{A_1, A_3, A_6, A_7, A_9, C_1\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_1, \mathcal{E}_7) = \{A_1, A_3, A_6, A_9, C_1\}$$
$$\mathsf{Prf\,Acc}^{\cup}(C_2, \mathcal{E}_{21}) = \{A_2, A_3, A_6, A_7, A_9, B_2, C_2\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_2, \mathcal{E}_{23}) = \{A_2, A_3, A_6, A_9, B_2, C_2\}$$
$$\mathsf{Prf\,Acc}^{\cup}(C_2, \mathcal{E}_{30}) = \{A_2, A_6, A_7, A_9, B_2, C_2\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_2, \mathcal{E}_{33}) = \{A_2, A_6, A_9, B_2, C_2\}$$
$$\mathsf{Prf\,Acc}^{\cup}(C_3, \mathcal{E}_{17}) = \{A_2, A_3, A_5, A_7, A_9, B_3, C_3\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_3, \mathcal{E}_{25}) = \{A_2, A_5, A_7, A_9, B_3, C_3\}$$
$$\mathsf{Prf\,Acc}^{\cup}(C_4, \mathcal{E}_{19}) = \{A_2, A_5, A_8, A_9, B_4, C_4\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_4, \mathcal{E}_{28}) = \{A_2, A_4, A_5, A_8, A_9, B_4, C_4, C_5\}$$
$$\mathsf{Prf\,Acc}^{\cup}(C_5, \mathcal{E}_9) = \{A_1, A_4, A_5, A_9, C_5\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_5, \mathcal{E}_{11}) = \{A_1, A_4, A_5, A_8, A_9, C_5\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_5, \mathcal{E}_{13}) = \{A_1, A_4, A_6, A_9, C_5\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_5, \mathcal{E}_{15}) = \{A_1, A_4, A_6, A_8, A_9, C_5\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_5, \mathcal{E}_{26}) = \{A_2, A_4, A_5, A_9, C_5\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_5, \mathcal{E}_{28}) = \{A_2, A_4, A_5, A_8, A_9, B_4, C_4, C_5\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_5, \mathcal{E}_{31}) = \{A_2, A_4, A_6, A_9, C_5\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_5, \mathcal{E}_{34}) = \{A_2, A_4, A_6, A_8, A_9, C_5\}$$
$$\mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_2) = \{A_1, A_5, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_4) = \{A_1, A_5, A_8, A_{10}, C_6\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_6) = \{A_1, A_6, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_8) = \{A_1, A_6, A_8, A_{10}, C_6\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{10}) = \{A_1, A_4, A_5, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{12}) = \{A_1, A_4, A_8, A_{10}, C_6\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{14}) = \{A_1, A_4, A_6, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{16}) = \{A_1, A_4, A_6, A_8, A_{10}, C_6\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{18}) = \{A_2, A_5, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{20}) = \{A_2, A_5, A_8, A_{10}, B_4, C_6\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{22}) = \{A_2, A_6, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{24}) = \{A_2, A_6, A_8, A_{10}, C_6\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{27}) = \{A_2, A_4, A_5, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{29}) = \{A_2, A_4, A_5, A_8, A_{10}, B_4, C_6\},$$
$$\mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{32}) = \{A_2, A_4, A_6, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}(C_6, \mathcal{E}_{35}) = \{A_2, A_4, A_6, A_8, A_{10}, C_6\}.$$

For $\preceq$-minimal explanations (with $\preceq \in \{\leq, \subseteq\}$), we use the notation $\mathsf{Prf\,Acc}^{\cup}_{\preceq}$, to denote the possible $\preceq$-minimal explanations from Section 6.1:

$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(C_1, \mathcal{E}_3) = \{A_1, A_3, A_5, A_9, C_1\} \qquad \mathsf{Prf\,Acc}^{\cup}_{\leq}(C_1, \mathcal{E}_7) = \{A_1, A_3, A_6, A_9, C_1\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(C_2, \mathcal{E}_{33}) = \{A_2, A_6, A_9, B_2, C_2\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(C_3, \mathcal{E}_{25}) = \{A_2, A_5, A_7, A_9, B_3, C_3\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(C_4, \mathcal{E}_{19}) = \{A_2, A_5, A_8, A_9, B_4, C_4\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(C_5, \mathcal{E}_9) = \{A_1, A_4, A_5, A_9, C_5\} \qquad \mathsf{Prf\,Acc}^{\cup}_{\leq}(C_5, \mathcal{E}_{13}) = \{A_1, A_4, A_6, A_9, C_5\},$$
$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(C_5, \mathcal{E}_{26}) = \{A_2, A_4, A_5, A_9, C_5\} \qquad \mathsf{Prf\,Acc}^{\cup}_{\leq}(C_5, \mathcal{E}_{31}) = \{A_2, A_4, A_6, A_9, C_5\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(C_6, \mathcal{E}_2) = \{A_1, A_5, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}_{\leq}(C_6, \mathcal{E}_6) = \{A_1, A_6, A_{10}, C_6\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(C_6, \mathcal{E}_{18}) = \{A_2, A_5, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}_{\leq}(C_6, \mathcal{E}_{22}) = \{A_2, A_6, A_{10}, C_6\}\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(C_6, \mathcal{E}_2) = \{\{A_1, A_5, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}_{\subseteq}(C_6, \mathcal{E}_6) = \{A_1, A_6, A_{10}, C_6\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(C_6, \mathcal{E}_{16}) = \{A_1, A_4, A_8, A_{10}, C_6\} \qquad \mathsf{Prf\,Acc}^{\cup}_{\subseteq}(C_6, \mathcal{E}_{18}) = \{A_2, A_5, A_{10}, C_6\}$$
$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(C_6, \mathcal{E}_{22}) = \{A_2, A_6, A_{10}, C_6\}\}.$$

For necessity and $\leq$-minimal sufficiency we obtain:

$$\mathsf{MinSuff}(C_1,\emptyset) = \{\{A_1,A_3,A_6,A_9,C_1\},\{A_1,A_3,A_7,A_9,C_1\},\{A_1,A_3,A_9,C_1\}\}$$
$$\mathsf{Nec}(C_1,\emptyset) = \{A_1,A_3,A_9,C_1\}$$
$$\mathsf{MinSuff}(C_2,\emptyset) = \{\{A_2,A_3,A_6,A_9,B_2,C_2\},\{A_2,A_6,A_9,B_2,C_2\}\}$$
$$\mathsf{Nec}(C_2,\emptyset) = \{A_2,A_6,A_9,B_2,C_2\}$$
$$\mathsf{MinSuff}(C_3,\emptyset) = \{\{A_2,A_3,A_5,A_7,A_9,B_3,C_3\},\{A_2,A_5,A_7,A_9,B_3,C_3\}\}$$
$$\mathsf{Nec}(C_3,\emptyset) = \{A_2,A_5,A_7,A_9,B_3,C_3\}$$
$$\mathsf{MinSuff}(C_4,\emptyset) = \{\{A_2,A_5,A_8,A_9,B_4,C_4\},\{A_5,A_8,A_9,B_4,C_4\}\}$$
$$\mathsf{Nec}(C_4,\emptyset) = \{A_5,A_8,A_9,C_4\}$$
$$\mathsf{MinSuff}(C_5,\emptyset) = \{\{A_4,A_9,C_5\}\}$$
$$\mathsf{Nec}(C_5,\emptyset) = \{A_4,A_9,C_5\}$$
$$\mathsf{MinSuff}(C_6,\emptyset) = \mathsf{Nec}(C_6,\emptyset) = \{A_{10},C_6\}.$$

We now turn to the explanations for the credulous acceptance of fraud respectively not fraud (i.e., f and $\neg f$). First for $\mathbb{F} = \mathsf{id}$:

$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_1) = \{\langle C_1,\{A_1,A_3,A_5,A_7,A_9,C_1\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_3) = \{\langle C_1,\{A_1,A_3,A_5,A_9,C_1\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_5) = \{\langle C_1,\{A_1,A_3,A_6,A_7,A_9,C_1\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_7) = \{\langle C_1,\{A_1,A_3,A_6,A_9,C_1\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_{17}) = \{\langle C_3,\{A_2,A_3,A_5,A_7,A_9,B_3,C_3\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_{21}) = \{\langle C_2,\{A_2,A_3,A_6,A_7,A_9,B_2,C_2\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_{23}) = \{\langle C_2,\{A_2,A_3,A_6,A_9,B_2,C_2\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_{25}) = \{\langle C_3,\{A_2,A_5,A_7,A_9,B_3,C_3\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_{30}) = \{\langle C_2,\{A_2,A_6,A_7,A_9,B_2,C_2\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(f,\mathcal{E}_{33}) = \{\langle C_2,\{A_2,A_6,A_9,B_2,C_2\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_2) = \{\langle C_6,\{A_1,A_5,A_{10},C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_4) = \{\langle C_6,\{A_1,A_5,A_8,A_{10},C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_6) = \{\langle C_6,\{A_1,A_6,A_{10},C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_8) = \{\langle C_6,\{A_1,A_6,A_8,A_{10},C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_9) = \{\langle C_5,\{A_1,A_4,A_5,A_9,C_5\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{10}) = \{\langle C_6,\{A_1,A_4,A_5,A_{10},C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{11}) = \{\langle C_5,\{A_1,A_4,A_5,A_8,A_9,C_5\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{12}) = \{\langle C_6,\{A_1,A_4,A_8,A_{10},C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{13}) = \{\langle C_5,\{A_1,A_4,A_6,A_9,C_5\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{14}) = \{\langle C_6,\{A_1,A_4,A_6,A_{10},C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{15}) = \{\langle C_5,\{A_1,A_4,A_6,A_8,A_9,C_5\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{16}) = \{\langle C_6,\{A_1,A_4,A_6,A_8,A_{10},C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{18}) = \{\langle C_6,\{A_2,A_5,A_{10},C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{19}) = \{\langle C_4,\{A_2,A_5,A_8,A_9,B_4,C_4\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{20}) = \{\langle C_6,\{A_2,A_5,A_8,A_{10},B_4,C_6\}\rangle\}$$
$$\mathsf{Prf}\,\mathsf{Acc}^{\cup}(\neg f,\mathcal{E}_{22}) = \{\langle C_6,\{A_2,A_6,A_{10},C_6\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{24}) = \{\langle C_6, \{A_2, A_6, A_8, A_{10}, C_6\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{26}) = \{\langle C_5, \{A_2, A_4, A_5, A_9, C_5\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{27}) = \{\langle C_6, \{A_2, A_4, A_5, A_{10}, C_6\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{28}) = \{\langle C_4, \{A_2, A_4, A_5, A_8, A_9, B_4, C_4, C_5\}\rangle, \langle C_5, \{A_2, A_4, A_5, A_8, A_9, B_4, C_4, C_5\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{29}) = \{\langle C_6, \{A_2, A_4, A_5, A_8, A_{10}, B_4, C_6\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{31}) = \{\langle C_5, \{A_2, A_4, A_6, A_9, C_5\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{32}) = \{\langle C_6, \{A_2, A_4, A_6, A_{10}, C_6\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{34}) = \{\langle C_5, \{A_2, A_4, A_6, A_8, A_9, C_5\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{35}) = \{\langle C_6, \{A_2, A_4, A_6, A_8, A_{10}, C_6\}\rangle\}$$

As discussed in Section 5.3, we can also choose $\mathbb{F} = \mathsf{Prem}$. The explanations for the credulous acceptance of f and $\neg f$ are then:

$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_1) = \{\langle\{cpd, fake, cd\}, \{cpd, fake, con, fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_3) = \{\langle\{cpd, fake, cd\}, \{cpd, fake, con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_5) = \{\langle\{cpd, fake, cd\}, \{cpd, fake, \neg con, fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_7) = \{\langle\{cpd, fake, cd\}, \{cpd, fake, \neg con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_{17}) = \{\langle\{\neg cpd, con, fcon, cd\}, \{\neg cpd, fake, con, fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_{21}) = \{\langle\{\neg cpd, \neg con, cd\}, \{\neg cpd, fake, \neg con, fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_{23}) = \{\langle\{\neg cpd, \neg con, cd\}, \{\neg cpd, fake, \neg con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_{25}) = \{\langle\{\neg cpd, con, fcon, cd\}, \{\neg cpd, con, fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_{30}) = \{\langle\{\neg cpd, \neg con, cd\}, \{\neg cpd, \neg con, fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(f, \mathcal{E}_{33}) = \{\langle\{\neg cpd, \neg con, cd\}, \{\neg cpd, \neg con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_2) = \{\langle\{\neg cd\}, \{cpd, con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_4) = \{\langle\{\neg cd\}, \{cpd, con, \neg fcon, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_6) = \{\langle\{\neg cd\}, \{cpd, \neg con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_8) = \{\langle\{\neg cd\}, \{cpd, \neg con, \neg fcon, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_9) = \{\langle\{\neg fake, cd\}, \{cpd, \neg fake, con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{10}) = \{\langle\{\neg cd\}, \{cpd, \neg fake, con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{11}) = \{\langle\{\neg fake, cd\}, \{cpd, \neg fake, con, \neg fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{12}) = \{\langle\{\neg cd\}, \{cpd, \neg fake, con, \neg fcon, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{13}) = \{\langle\{\neg fake, cd\}, \{cpd, \neg fake, \neg con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{14}) = \{tuple\{\neg cd\}, \{cpd, \neg fake, \neg con, \neg cd\}\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{15}) = \{\langle\{\neg fake, cd\}, \{cpd, \neg fake, \neg con, \neg fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{16}) = \{\langle\{\neg cd\}, \{cpd, \neg fake, \neg con, \neg fcon, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{18}) = \{\langle\{\neg cd\}, \{\neg cpd, con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{19}) = \{\langle\{\neg cpd, con, \neg fcon, cd\}, \{\neg cpd, con, \neg fcon, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{20}) = \{\langle\{\neg cd\}, \{\neg cpd, con, \neg fcon, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{22}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{24}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg con, \neg fcon, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{26}) = \{\langle\{\neg fake, cd\}, \{\neg cpd, \neg fake, con, cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{27}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg fake, con, \neg cd\}\rangle\}$$
$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{28}) = \{\langle\{\neg cpd, con, \neg fcon, cd\}, \{\neg cpd, \neg fake, con, \neg fcon, cd\}\rangle,$$

$$\langle\{\neg fake, cd\}, \{\neg cpd, \neg fake, con, \neg fcon, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{29}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg fake, con, \neg fcon, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{31}) = \{\langle\{\neg fake, cd\}, \{\neg cpd, \neg fake, \neg con, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{32}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg fake, \neg con, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{34}) = \{\langle\{\neg fake, cd\}, \{\neg cpd, \neg fake, \neg con, \neg fcon, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}(\neg f, \mathcal{E}_{35}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg fake, \neg con, \neg fcon, \neg cd\}\rangle\}$$

Like for argument explanations, we will first look at the minimal explanations, again using the notation $\mathsf{Prf\,Acc}^{\cup}_{\leq}$ to denote the possible $\leq$-minimal explanations:

$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(f, \mathcal{E}_{33}) = \langle\{\neg cpd, \neg con, cd\}, \{\neg cpd, \neg con, cd\}\rangle$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(f, \mathcal{E}_{3}) = \{\langle\{cpd, fake, cd\}, \{cpd, fake, con, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(f, \mathcal{E}_{7}) = \{\langle\{cpd, fake, cd\}, \{cpd, fake, \neg con, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(f, \mathcal{E}_{23}) = \{\langle\{\neg cpd, \neg con, cd\}, \{\neg cpd, fake, \neg con, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(f, \mathcal{E}_{25}) = \{\langle\{\neg cpd, con, fcon, cd\}, \{\neg cpd, con, fcon, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(f, \mathcal{E}_{33}) = \langle\{\neg cpd, \neg con, cd\}, \{\neg cpd, \neg con, cd\}\rangle$$

$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(\neg f, \mathcal{E}_{2}) = \{\langle\{\neg cd\}, \{cpd, con, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(\neg f, \mathcal{E}_{6}) = \{\langle\{\neg cd\}, \{cpd, \neg con, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(\neg f, \mathcal{E}_{18}) = \{\langle\{\neg cd\}, \{\neg cpd, con, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\leq}(\neg f, \mathcal{E}_{22}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg con, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{2}) = \{\langle\{\neg cd\}, \{cpd, con, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{6}) = \{\langle\{\neg cd\}, \{cpd, \neg con, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{9}) = \{\langle\{\neg fake, cd\}, \{cpd, \neg fake, con, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{13}) = \{\langle\{\neg fake, cd\}, \{cpd, \neg fake, \neg con, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{18}) = \{\langle\{\neg cd\}, \{\neg cpd, con, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{19}) = \{\langle\{\neg cpd, con, \neg fcon, cd\}, \{\neg cpd, con, \neg fcon, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{22}) = \{\langle\{\neg cd\}, \{\neg cpd, \neg con, \neg cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{26}) = \{\langle\{\neg fake, cd\}, \{\neg cpd, \neg fake, con, cd\}\rangle\}$$

$$\mathsf{Prf\,Acc}^{\cup}_{\subseteq}(\neg f, \mathcal{E}_{31}) = \{\langle\{\neg fake, cd\}, \{\neg cpd, \neg fake, \neg con, cd\}\rangle\}$$

Finally, we turn to necessity and $\leq$-minimal sufficiency:

$$\text{Minimal sufficiency:} \quad \mathsf{Suff\,Acc}^{\cup}_{\leq}(f, \emptyset) = \{\langle\{cpd, fake, cd\}, \{cpd, fake, cd\}\rangle, \langle\{\neg cpd, \neg con, cd\}, \{\neg cpd, \neg con, cd\}\rangle\}$$

$$\text{Necessity:} \quad \mathsf{NecAcc}^{\cup}(f, \emptyset) = \langle\{cd\}, \{cd\}\rangle$$

$$\text{Minimal sufficiency:} \quad \mathsf{Suff\,Acc}^{\cup}_{\leq}(\neg f, \emptyset) = \{\langle\{\neg cd\}, \{\neg cd\}\rangle\}$$

$$\text{Necessity:} \quad \mathsf{NecAcc}^{\cup}(\neg f, \emptyset) = \emptyset.$$

References

[1] C. Antaki, I. Leudar, Explaining in conversation: towards an argument model, Eur. J. Soc. Psychol. 22 (2) (1992) 181–194.

[2] O. Arieli, A. Borg, J. Heyninck, C. Straßer, Logic-based approaches to formal argumentation, in: D. Gabbay, M. Giacomin, G.R. Simari, M. Thimm (Eds.), Handbook of Formal Argumentation, vol. 2, College Publications, 2021, pp. 719–850.

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

[4] K. Atkinson, T. Bench-Capon, D. Bollegala, Explanation in AI and law: past, present and future, Artif. Intell. 289 (2020) 103387.

[5] P. Baroni, M. Caminada, M. Giacomin, Abstract argumentation frameworks and their semantics, in: P. Baroni, D. Gabay, M. Giacomin, L. van der Torre (Eds.), Handbook of Formal Argumentation, College Publications, 2018, pp. 159–236.

[6] P. Baroni, A. Rago, F. Toni, From fine-grained properties to broad principles for gradual argumentation: a principled spectrum, Int. J. Approx. Reason. 105 (2019) 252–286.

[7] R. Baumann, S. Doutre, J.-G. Mailly, J. Wallner, Enforcement in formal argumentation, in: D.M. Gabbay, M. Giacomin, G.R. Simari, M. Thimm (Eds.), Handbook of Formal Argumentation, vol. 2, College Publications, 2021, pp. 445–510.

[8] T. Bench-Capon, P. Dunne, Argumentation in artificial intelligence, Artif. Intell. 171 (10) (2007) 619–641.
[9] P. Besnard, A. Garcia, A. Hunter, S. Modgil, H. Prakken, G. Simari, F. Toni, Introduction to structured argumentation, Argument Comput. 5 (1) (2014) 1–4.
[10] P. Besnard, S. Doutre, T. Duchatelle, M. Lagasquie-Schiex, Explaining semantics and extension membership in abstract argumentation, Intell. Syst. Appl. 16 (2022) 200118.
[11] F. Bex, D. Walton, Combining explanation and argumentation in dialogue, Argument Comput. 7 (1) (2016) 55–68.
[12] F. Bex, B. Testerink, J. Peters, AI for online criminal complaints: from natural dialogues to structured scenarios, in: Workshop Proceedings of Artificial Intelligence for Justice at ECAI 2016, 2016, pp. 22–29.
[13] E. Black, N. Maudet, S. Parsons, Argumentation-based dialogue, in: D. Gabbay, M. Giacomin, G.R. Simari, M. Thimm (Eds.), Handbook of Formal Argumentation, vol. 2, College Publications, 2021, pp. 511–575.
[14] G. Boella, S. Kaci, L. van der Torre, Dynamics in argumentation with single extensions: abstraction principles and the grounded extension, in: C. Sossai, G. Chemello (Eds.), Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'09), in: Lecture Notes in Computer Science, vol. 5590, Springer, 2009, pp. 107–118.
[15] A. Bondarenko, P.M. Dung, R. Kowalski, F. Toni, An abstract, argumentation-theoretic approach to default reasoning, Artif. Intell. 93 (1) (1997) 63–101.
[16] A. Borg, F. Bex, A basic framework for explanations in argumentation, IEEE Intell. Syst. 36 (2) (2021) 25–35.
[17] A. Borg, F. Bex, Necessary and sufficient explanations for argumentation-based conclusions, in: J. Vejnarová, N. Wilson (Eds.), Proceedings of the 16th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'21), in: Lecture Notes in Computer Science, vol. 12897, Springer, 2021, pp. 45–58.
[18] A. Borg, F. Bex, Enforcing sets of formulas in structured argumentation, in: M. Bienvenu, G. Lakemeyer, E. Erdem (Eds.), Proceedings of the 18th International Conference on Principles of Knowledge Representation and Reasoning (KR'21), IJCAI Organization, 2021, pp. 130–140.
[19] M. Caminada, P.E. Dunne, Minimal strong admissibility: a complexity analysis, in: H. Prakken, S. Bistarelli, F. Santini, C. Taticchi (Eds.), Proceedings of the 8th International Conference on Computational Models of Argument (COMMA'20), in: Frontiers in Artificial Intelligence and Applications, vol. 326, IOS Press, 2020, pp. 135–146.
[20] M. Caminada, Y. Wu, On the limitations of abstract argumentation, in: Proceedings of the 23rd Benelux Conference on Artificial Intelligence (BNAIC'11), 2011.
[21] O. Cocarascu, A. Stylianou, K. Cyras, F. Toni, Data-empowered argumentation for dialectically explainable predictions, in: G.D. Giacomo, A. Catalá, B. Dilkina, M. Milano, S. Barro, A. Bugarín, J. Lang (Eds.), Proceedings of the 24th European Conference on Artificial Intelligence (ECAI'20), in: Frontiers in Artificial Intelligence and Applications, vol. 325, IOS Press, 2020, pp. 2449–2456.
[22] L. Console, P. Torasso, A spectrum of logical definitions of model-based diagnosis 1, Comput. Intell. 7 (3) (1991) 133–141.
[23] K. Čyras, D. Birch, Y. Guo, F. Toni, R. Dulay, S. Turvey, D. Greenberg, T. Hapuarachchi, Explanations by arbitrated argumentative dispute, Expert Syst. Appl. 127 (2019) 141–156.
[24] K. Čyras, A. Rago, E. Albini, P. Baroni, F. Toni, Argumentative XAI: a survey, in: Z. Zhou (Ed.), Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI'21), 2021, pp. 4392–4399, ijcai.org.
[25] A. d'Avila Garcez, L. Lamb, Neurosymbolic AI: the 3rd wave, arXiv:2012.05876 [abs], 2020.
[26] A. Dejl, P. He, P. Mangal, H. Mohsin, B. Surdu, E. Voinea, E. Albini, P. Lertvittayakumjorn, A. Rago, F. Toni, Argflow: a toolkit for deep argumentative explanations for neural networks, in: Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS'21), International Foundation for Autonomous Agents and Multiagent Systems, 2021, pp. 1761–1763.
[27] J. Delobelle, S. Villata, Interpretability of gradual semantics in abstract argumentation, in: G. Kern-Isberner, Z. Ognjanovic (Eds.), Proceedings of the 15th European Conference Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'19), in: Lecture Notes in Computer Science, vol. 11726, Springer, 2019, pp. 27–38.
[28] A. Dhurandhar, P. Chen, R. Luss, C. Tu, P. Ting, K. Shanmugam, P. Das, Explanations based on the missing: towards contrastive explanations with pertinent negatives, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS'18), 2018, pp. 590–601.
[29] S. Doutre, J. Mengin, On sceptical versus credulous acceptance for abstract argument systems, in: J.J. Alferes, J.A. Leite (Eds.), Proceedings of the 9th European Conference on Logics in Artificial Intelligence (JELIA'04), in: Lecture Notes in Computer Science, vol. 3229, Springer, 2004, pp. 462–473.
[30] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (2) (1995) 321–357.
[31] P.M. Dung, R. Kowalski, F. Toni, Assumption-based argumentation, in: G.R. Simari, I. Rahwan (Eds.), Argumentation in Artificial Intelligence, Springer, 2009, pp. 199–218.
[32] W. Dvořák, P.E. Dunne, Computational problems in formal argumentation and their complexity, in: P. Baroni, D. Gabay, M. Giacomin, L. van der Torre (Eds.), Handbook of Formal Argumentation, College Publications, 2018, pp. 631–688.
[33] X. Fan, F. Toni, On computing explanations in argumentation, in: B. Bonet, S. Koenig (Eds.), Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI'15), AAAI Press, 2015, pp. 1496–1502.
[34] X. Fan, F. Toni, On explanations for non-acceptable arguments, in: E. Black, S. Modgil, N. Oren (Eds.), Proceedings of the 3rd International Workshop on Theory and Applications of Formal Argumentation, (TAFA'15), in: Lecture Notes in Computer Science, vol. 9524, Springer, 2015, pp. 112–127.
[35] A. García, G. Simari, Defeasible logic programming: an argumentative approach, Theory Pract. Log. Program. 4 (1&2) (2004) 95–138.
[36] A. García, C. Chesñevar, N. Rotstein, G. Simari, Formalizing dialectical explanation support for argument-based reasoning in knowledge-based systems, Expert Syst. Appl. 40 (8) (2013) 3233–3247.
[37] D. Hilton, Conversational processes and causal explanation, Psychol. Bull. 107 (1) (1990) 65–81.
[38] T. Kampik, K. Čyras, J. Ruiz Alarcón, Change in quantitative bipolar argumentation: sufficient, necessary, and counterfactual explanations, Int. J. Approx. Reason. 164 (2024) 109066.
[39] D. Lewis, Causal explanation, Philos. Pap. 2 (1986) 214–240.
[40] B. Liao, L. van der Torre, Explanation semantics for abstract argumentation, in: H. Prakken, S. Bistarelli, F. Santini, C. Taticchi (Eds.), Proceedings of the 8th International Conference on Computational Models of Argument (COMMA'20), in: Frontiers in Artificial Intelligence and Applications, vol. 326, IOS Press, 2020, pp. 271–282.
[41] F. Lin, On strongest necessary and weakest sufficient conditions, Artif. Intell. 128 (1) (2001) 143–159.
[42] P. Lipton, Contrastive explanation, R. Inst. Philos. Suppl. 27 (1990) 247–266.
[43] T. Lombrozo, Causal-explanatory pluralism: how intentions, functions, and mechanisms influence causal ascriptions, Cogn. Psychol. 61 (4) (2010) 303–332.
[44] G. Marcus, The next decade in AI: four steps towards robust artificial intelligence, CoRR, arXiv:2002.06177 [abs], 2020, https://arxiv.org/abs/2002.06177.
[45] G. Marcus, E. Davis, Rebooting AI: Building Artificial Intelligence We Can Trust, Vintage Books, 2019.
[46] T. Miller, Explanation in artificial intelligence: insights from the social sciences, Artif. Intell. 267 (2019) 1–38.
[47] S. Modgil, M. Caminada, Proof theories and algorithms for abstract argumentation frameworks, in: G.R. Simari, I. Rahwan (Eds.), Argumentation in Artificial Intelligence, Springer, 2009, pp. 105–129.
[48] S. Modgil, H. Prakken, Resolutions in structured argumentation, in: B. Verheij, S. Szeider, S. Woltran (Eds.), Proceedings of the 4th International Conference on Computational Models of Argument (COMMA'12), in: Frontiers in Artificial Intelligence and Applications, vol. 245, IOS Press, 2012, pp. 310–321.
[49] S. Modgil, H. Prakken, A general account of argumentation with preferences, Artif. Intell. 195 (2013) 361–397.

[50] S. Modgil, H. Prakken, The ASPIC+ framework for structured argumentation: a tutorial, Argument Comput. 5 (1) (2014) 31–62.
[51] S. Modgil, H. Prakken, Abstract rule-based argumentation, in: P. Baroni, D. Gabay, M. Giacomin, L. van der Torre (Eds.), Handbook of Formal Argumentation, College Publications, 2018, pp. 287–364s.
[52] A. Niskanen, M. Järvisalo, Smallest explanations and diagnoses of rejection in abstract argumentation, in: D. Calvanese, E. Erdem, M. Thielscher (Eds.), Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR'20), 2020, pp. 667–671.
[53] D. Odekerken, F. Bex, Towards transparent human-in-the-loop classification of fraudulent web shops, in: S. Villata, J. Harašta, P. Křemen (Eds.), Proceedings of the 33rd International Conference on Legal Knowledge and Information Systems (JURIX'20), in: Frontiers in Artificial Intelligence and Applications, vol. 334, IOS Press, 2020, pp. 239–242.
[54] D. Odekerken, F. Bex, A. Borg, B. Testerink, Approximating stability for applied argument-based inquiry, Intell. Syst. Appl. 16 (2022) 200110.
[55] N. Potyka, X. Yin, F. Toni, Explaining random forests using bipolar argumentation and Markov networks, in: B. Williams, Y. Chen, J. Neville (Eds.), Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI Press, 2023, pp. 9453–9460.
[56] H. Prakken, An abstract framework for argumentation with structured arguments, Argument Comput. 1 (2) (2010) 93–124.
[57] H. Prakken, Historical overview of formal argumentation, in: P. Baroni, D. Gabay, M. Giacomin, L. van der Torre (Eds.), Handbook of Formal Argumentation, College Publications, 2018, pp. 75–143.
[58] H. Prakken, M. de Winter, Abstraction in argumentation: necessary but dangerous, in: S. Modgil, K. Budzynska, J. Lawrence (Eds.), Proceedings of the 7th International Conference on Computation Models of Argument (COMMA'18), in: Frontiers in Artificial Intelligence and Applications, vol. 305, IOS Press, 2018, pp. 85–96.
[59] A. Rago, O. Cocarascu, C. Bechlivanidis, D. Lagnado, F. Toni, Argumentative explanations for interactive recommendations, Artif. Intell. 296 (2021) 103506.
[60] A. Rotolo, G. Sartor, Argumentation and explanation in the law, Front. Artif. Intell. 6 (2023).
[61] Z. Saribatur, J. Wallner, S. Woltran, Explaining non-acceptability in abstract argumentation, in: Proceedings of the 24th European Conference on Artificial Intelligence (ECAI'20), in: Frontiers in Artificial Intelligence and Applications, vol. 325, IOS Press, 2020, pp. 881–888.
[62] P. Thagard, Explanatory coherence, Behav. Brain Sci. 12 (3) (1989) 435–502.
[63] M. Ulbricht, J.P. Wallner, Strong explanations in abstract argumentation, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 6496–6504.
[64] A. Vassiliades, N. Bassiliades, T. Patkos, Argumentation and explainable artificial intelligence: a survey, Knowl. Eng. Rev. 36 (2021) e5.
[65] D.S. Watson, L. Gultchin, A. Taly, L. Floridi, Local explanations via necessity and sufficiency: unifying theory and practice, CoRR, arXiv:2103.14651 [abs], 2021, https://arxiv.org/abs/2103.14651.
[66] J. Woodward, Sensitive and insensitive causation, Philos. Rev. 115 (1) (2006) 1–50.
[67] Z. Zeng, C. Miao, C. Leung, Z. Shen, J.J. Chin, Computing argumentative explanations in bipolar argumentation frameworks, in: The 33rd AAAI Conference on Artificial Intelligence (AAAI'19), AAAI Press, 2019, pp. 10079–10080.