# On Serialisability for Argumentative Explanations

Lars Bengel

Artificial Intelligence Group, University of Hagen, Germany

**Abstract**

We consider the recently proposed notion of *serialisability* of semantics for abstract argumentation frameworks. This notion describes a method for the serialised non-deterministic construction of extensions through iterative addition of non-empty minimal admissible sets. This paper provides an overview on serialisability and proposes some promising applications of this concept in the context of *explainable AI*. In particular, we outline how serialisability could be employed to provide more structured explanations. We also discuss how serialisability could be utilised in discussion games.

## 1 Introduction

Since the introduction of *abstract argumentation* by Dung in his seminal paper [Dung, 1995], the research field has received great attention in the literature. The goal of abstract argumentation is modelling real-world exchange of arguments and using this model for reasoning. An important concept in argumentation is *admissibility* which characterises sets of arguments (extensions) that are conflict-free and defend itself against all attackers.

In recent years, the need for human understandable explanations for AI models has grown stronger and as a result of that the field of *eXplainable AI* (XAI) [Adadi and Berrada, 2018] has seen many new proposals. One very natural approach to explanations is in fact formal argumentation [Antaki and Leudar, 1992] and there already exist many approaches to XAI that utilise argumentation [Čyras et al., 2021, Ulbricht and Wallner, 2021].

In [Xu and Cayrol, 2016] the notion of *initial sets*, i. e., non-empty minimal admissible sets, has been introduced. The intuition behind initial sets is that they each solve an atomic conflict of the argumentation framework. Built on that idea, the concept of *serialisability* [Thimm, 2022] has emerged as a novel means for constructing extensions, with initial sets as the building blocks.

In this work, we take a closer look at serialisability. We will recall some work that we have already done on this topic [Bengel and Thimm, 2022] and also provide an outlook on some applications of serialisability, in particular in the domain of explainability. We consider how serialisability could be used to provide structured explanations for the acceptance of an extension. The serialised construction of an extension essentially gives us a decomposition of the extension into a series of initial sets of the respective reducts. This helps by providing a clearer view on why the extension is acceptable and allows for the natural construction of an explanation for its acceptance.

In interesting and related field are the *discussion games* [Caminada, 2015, Caminada, 2018]. In these games, two agents try to win a discussion by taking turns providing arguments that refute those of the opponent. The idea is that the presence of a winning strategy for a semantics $\sigma$ guarantees the existence of a $\sigma$-extension. Caminada also shows that there are different types of discussion games, corresponding to different semantics. The discussions induced by these games can be seen as a kind of explanation for the acceptance of arguments. In the following, we will also explore how serialisability could be applied in the context of the discussion games.

The remainder of this work is structured as follows. In Section 2, we introduce the necessary background on argumentation and serialisability. Following that, in Section 3 we recall some of the work that we have already done on the topic and discuss in more detail the above mentioned potential applications of serialisability. Section 4 concludes the paper.

# 2 Method

We consider the *abstract argumentation frameworks* as introduced by [Dung, 1995]. Let $\mathfrak{A}$ denote a universal set of arguments. An argumentation framework (AF) is represented as a graph whose nodes are arguments and the directed edges are conflicts between them, i. e., an edge between two arguments $a$ and $b$ means that $a$ attacks $b$.

**Definition 1.** An *abstract argumentation framework* $\mathsf{F}$ is a tuple $\mathsf{F} = (\mathsf{A}, \mathsf{R})$ where $\mathsf{A} \subseteq \mathfrak{A}$ is a finite set of arguments and $\mathsf{R}$ is a relation $\mathsf{R} \subseteq \mathsf{A} \times \mathsf{A}$.

With $\mathfrak{AF}$ we denote the set of all argumentation frameworks. For a set $X \subseteq \mathsf{A}$, we denote by $\mathsf{F}|_X = (X, \mathsf{R} \cap (X \times X))$ the projection of $\mathsf{F}$ on $X$. For a set $S \subseteq \mathsf{A}$ we define $S^+ = \{a \in \mathsf{A} \mid \exists b \in S : b\mathsf{R}a\}$ and $S^- = \{a \in \mathsf{A} \mid \exists b \in S : a\mathsf{R}b\}$.

We say that a set $S \subseteq \mathsf{A}$ is *conflict-free* if for all $a, b \in S$ we do not have that $a\mathsf{R}b$. A set $S$ *defends* an argument $b \in \mathsf{A}$ if for all $a$ with $a\mathsf{R}b$ there is an argument $c \in S$ such that $c\mathsf{R}a$. A conflict-free set $S$ is called *admissible* if $S$ defends all $a \in S$. Let $\mathsf{adm}(\mathsf{F})$ denote the set of admissible sets of $\mathsf{F}$.

We define different semantics by imposing constraints on admissible sets [Baroni et al., 2018].

**Definition 2.** Let $\mathsf{F} = (\mathsf{A}, \mathsf{R})$ be an argumentation framework. An admissible set $E$

- is a *complete* ($\mathsf{co}$) extension iff for all $a \in \mathsf{A}$, if $E$ defends $a$ then $a \in E$,
- is a *grounded* ($\mathsf{gr}$) extension iff $E$ is complete and minimally so,
- is a *preferred* ($\mathsf{pr}$) extension iff $E$ is maximal.

All statements on minimality/maximality are with respect to set inclusion. For $\sigma \in \{\mathsf{co}, \mathsf{gr}, \mathsf{st}, \mathsf{pr}\}$ let $\sigma(\mathsf{F})$ denote the set of $\sigma$-extensions of $\mathsf{F}$.

In [Xu and Cayrol, 2016], the authors introduce the notion of *initial sets*. They are defined as the non-empty, minimal admissible sets of an argumentation framework.

**Definition 3.** For $\mathsf{F} = (\mathsf{A}, \mathsf{R})$, a set $S \subseteq \mathsf{A}$ with $S \neq \emptyset$ is called an *initial set* if $S$ is admissible and there is no admissible $S' \subsetneq S$ with $S' \neq \emptyset$. Let $\mathsf{IS}(\mathsf{F})$ denote the set of initial sets of $\mathsf{F}$.

The intuition behind initials sets is that they each solve an atomic conflict of the argumentation framework. Due to the minimality, an initial set $S$ contains only those arguments that actually contribute to the defense of $S$. In other words, all elements of $S$ are relevant to the conflict that $S$ solves. The work [Thimm, 2022] introduces a useful distinction between three different types of initial sets.

**Definition 4.** For $\mathsf{F} = (\mathsf{A}, \mathsf{R})$, $S \in \mathsf{IS}(\mathsf{F})$, we say that

1. $S$ is *unattacked* iff $S^- = \emptyset$,
2. $S$ is *unchallenged* iff $S^- \neq \emptyset$ and there is no $S' \in \mathsf{IS}(\mathsf{F})$ with $S'\mathsf{R}S$,
3. $S$ is *challenged* iff there is $S' \in \mathsf{IS}(\mathsf{F})$ with $S'\mathsf{R}S$.

In the following, we will denote with $\mathsf{IS}^{\not\leftarrow}(\mathsf{F})$, $\mathsf{IS}^{\not\leftrightarrow}(\mathsf{F})$ and $\mathsf{IS}^{\leftrightarrow}(\mathsf{F})$ the set of unattacked, unchallenged, and challenged initial sets, respectively.

Based on the initial sets, [Thimm, 2022] introduced the notion of *serialisability*. This is a new approach for iteratively constructing the extensions of an admissible-based semantics via initial sets. For that, we first recall the notion of the *reduct* from [Baumann et al., 2020b].

**Definition 5.** For $\mathsf{F} = (\mathsf{A}, \mathsf{R})$ and $S \subseteq \mathsf{A}$, the $S$-reduct $\mathsf{F}^S$ is defined via $\mathsf{F}^S = \mathsf{F}|_{A \setminus (S \cup S^+)}$.

Intuitively, the construction process works as follows: First, we solve an atomic conflict in $\mathsf{F}$ by selecting some initial set $S$. Afterwards, we move to the reduct $\mathsf{F}^S$ which may reveal new conflicts and therefore different initial sets. We continue this process until some termination criterion is satisfied.

Formally, this method for constructing extensions is represented as a transition system. For the step

of selecting an initial set (for the transition) we need a *selection function* $\alpha$. Additionally, we also require a criterion $\beta$ for determining if the construction of an extension is finished. The following concepts have been defined for this purpose.

**Definition 6.** A *state* $T$ is a tuple $T = (\mathsf{F}, S)$ with $\mathsf{F} = (\mathsf{A}, \mathsf{R})$ and $S \subseteq \mathsf{A}$.

**Definition 7.** A *selection function* $\alpha$ is any function $\alpha : 2^{2^{\mathfrak{A}}} \times 2^{2^{\mathfrak{A}}} \times 2^{2^{\mathfrak{A}}} \to 2^{2^{\mathfrak{A}}}$ with $\alpha(X, Y, Z) \subseteq X \cup Y \cup Z$ for all $X, Y, Z \subseteq 2^{\mathfrak{A}}$.

The selection function will be applied as $\alpha(\mathsf{IS}^{\not\leftarrow}(\mathsf{F}), \mathsf{IS}^{\not\rightarrow}(\mathsf{F}), \mathsf{IS}^{\leftrightarrow}(\mathsf{F}))$ for some argumentation framework $\mathsf{F}$. So $\alpha$ selects a subset of the initial sets that are eligible to be selected in the construction.

**Definition 8.** A *termination function* $\beta$ is any function $\beta : \mathfrak{AF} \times 2^{\mathfrak{A}} \to \{0, 1\}$.

The termination function $\beta$ is used to indicate when the construction of an extension is finished (this will be the case if $\beta(\mathsf{F}, S) = 1$).

The transition rule, for some selection function $\alpha$, is defined as follows:

$$(\mathsf{F}, S) \xrightarrow{S' \in \alpha(\mathsf{IS}^{\not\leftarrow}(\mathsf{F}), \mathsf{IS}^{\not\rightarrow}(\mathsf{F}), \mathsf{IS}^{\leftrightarrow}(\mathsf{F}))} (\mathsf{F}^{S'}, S \cup S').$$

If $(\mathsf{F}', S')$ can be reached from $(\mathsf{F}, S)$ via a finite number of steps (including no steps at all) with the above rule and also satisfies some termination criterion of $\beta$ we write $(\mathsf{F}, S) \rightsquigarrow^{\alpha,\beta} (\mathsf{F}', S')$. Given a concrete instance of $\alpha$ and $\beta$, let $\mathcal{E}^{\alpha,\beta}(\mathsf{F})$ be the set of all $S$ with $(\mathsf{F}, \emptyset) \rightsquigarrow^{\alpha,\beta} (\mathsf{F}', S)$ (for some $\mathsf{F}'$).

**Definition 9.** A semantics $\sigma$ is *serialisable* if there exists a selection function $\alpha_\sigma$ and a termination function $\beta_\sigma$ with $\sigma(\mathsf{F}) = \mathcal{E}^{\alpha,\beta}(\mathsf{F})$ for all $\mathsf{F}$.

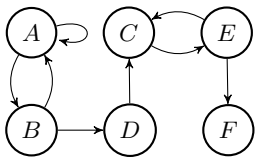


Figure 1: An AF with the initial sets: $\{B\}$ and $\{E\}$.

**Example 1.** Consider the preferred semantics, serialisable via $\alpha_{ad}(X, Y, Z) = X \cup Y \cup Z$ and

$$\beta_{pr}(\mathsf{F}, S) = \begin{cases} 1 & \text{if } \mathsf{IS}(\mathsf{F}) = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

We compute the preferred extensions of the AF $\mathsf{F}$ in Figure 1. We start the transition system in the state $(\mathsf{F}, \emptyset)$ and are allowed to select both $\{B\}$ and $\{F\}$. Assume we select $\{B\}$, thus we move to the state $(\mathsf{F}^{\{B\}}, \{B\})$. Now, in this state $\alpha_{ad}$ returns $\{\{C\}$ and $\{E\}\}$. Selecting $\{C\}$ then leads to the state $(\mathsf{F}^{\{B,C\}}, \{B, C\})$. Finally, only $\{F\}$ can be selected and we obtain the preferred extension $\{B, C, F\}$ of $\mathsf{F}$. Alternatively, selecting $\{E\}$ first leads to the only other preferred extension $\{B, E\}$.

Most admissible-based semantics are serialisable, namely admissible, strong admissible, complete, grounded, preferred and stable semantics [Thimm, 2022].

# 3 Discussion

We start with recalling some of the results already achieved in this area. In [Bengel and Thimm, 2022], we investigated the serialisability principle in more detail. In particular, we considered its relationship to other principles from the literature. Due to space limitations, we refer to [Baroni et al., 2005, Baroni and Giacomin, 2007, Baumann et al., 2020a] for the definitions of the mentioned principles. In our investigation, we found that every serialisable semantics also satisfies conflict-freeness, admissibility and modularization. On the other hand, serialisability does not imply directionality or SCC-recursiveness and vice versa.

We also introduced the property of $\alpha\beta$-closure for serialisable semantics, which is satisfied if every path of the corresponding transition system terminates, i. e., leads to some extension. Interestingly, this leads to the following connection.

**Theorem 1.** *If a semantics $\sigma$ is serialisable via $\alpha_\sigma$ and $\beta_\sigma$ and is $\alpha_\sigma\beta_\sigma$-closed, then $\sigma$ satisfies directionality.*

Furthermore, in [Bengel and Thimm, 2022] we also took a closer look at the unchallenged semantics from [Thimm, 2022] defined via the selection function $\alpha_{uc}(X, Y, Z) = X \cup Y$ and the termination function

$$\beta_{uc}(\mathsf{F}, S) = \begin{cases} 1 & \text{if } \mathsf{IS}^{\not\leftarrow}(\mathsf{F}) \cup \mathsf{IS}^{\not\rightarrow}(\mathsf{F}) = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

**Example 2.** Consider the AF $\mathsf{F}$ in Figure 1. Both $\{B\}$ and $\{E\}$ are unchallenged. If we select $\{B\}$ in the first step, then in the reduct $\mathsf{F}^{\{B\}}$ we have that $\{C\}$ and $\{E\}$ are now challenged initial sets. Thus $\beta_{uc}(\mathsf{F}^{\{B\}}, \{B\}) = 1$ and $\{B\}$ is an unchallenged extension of $\mathsf{F}$. However, if we select $\{E\}$ first, then $\{B\}$ can still be selected in the state $(\mathsf{F}^{\{E\}}, \{E\})$ and we reach the only other unchallenged extension $\{B, E\}$.

Essentially, this semantics amounts to exhaustively adding unattacked and unchallenged initial sets. The unchallenged semantics is $\alpha_{uc}, \beta_{uc}$-closed and thus satisfies directionality. It also satisfies reinstatement, but it does for example not satisfy SCC-recursiveness and I-maximality. We have also analysed the computational complexity of the relevant reasoning tasks.

We now turn to some applications of serialisability that we plan to explore in the future.

Using an argumentative approach to provide explanations is very natural [Antaki and Leudar, 1992]. An extension $E$ contains all relevant arguments to justify its own acceptance. However, a larger AF may consist of many different conflicts and for each only a subset of $E$ is of relevance. This is where serialisable semantics can be very useful. With the serialised construction of an extension $E$ we additionally obtain a decomposition of $E$ into a series of initial sets $S_1, ..., S_n$ of the respective reducts. Each of these initial sets solves an atomic conflict of the AF and only contains the arguments relevant to that conflict. Not only does this decompose an extension into initial sets, it also gives us information about their order and how some initial sets are only revealed once other conflicts are addressed. This information could, for example, be utilised to generate better, more structured explanations for an extension. In particular, in my PhD thesis, I want to investigate the relation to other recent approaches for acceptance explanations from the literature, such as related admissibility [Fan and Toni, 2015] or explanation schemes [Baumann and Ulbricht, 2021].

Another interesting application of serialisability could be the discussion games [Caminada, 2018]. In the course of such a game, each agent essentially iteratively constructs his own extension by adding one argument each round. This is somewhat similar to how the serialised construction works, where we select initial sets instead of individual arguments. Thus, it would be interesting to explore the relation between both concepts. There exist different discussion games for the different semantics. Notably, the notion of serialisability allows us to define completely new semantics by defining only a selection and a termination function. For example, confronted with some argument by our opponent, we may want to reply with the strongest refutation instead of simply providing any counterargument. For this purpose, we can define a selection function with some heuristic that evaluates the strength of initial sets (or arguments) and only allows us to select the strongest. This way, we could construct a strong, discussion-like explanation for the acceptance of our starting argument.

# 4 Conclusion

In this work we discussed the recently introduced notion of serialisability for argumentation semantics that provides a non-deterministic procedure for constructing extensions. We looked at some results relating serialisability to other principles from the literature. We also considered one instantiation, the unchallenged semantics, and its properties. For my PhD thesis, we outlined that the decomposition of an extension into initial sets could be used to provide more structured explanations. Furthermore, we also discussed the planned investigation of the similarities to discussion games [Caminada, 2015] and utilising serialisability for them.

**Acknowledgements** The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant 375588274).

## References


[Adadi and Berrada, 2018] Adadi, A. and Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, 6:52138–52160.

[Antaki and Leudar, 1992] Antaki, C. and Leudar, I. (1992). Explaining in conversation: Towards an argument model. *European Journal of Social Psychology*, 22(2):181–194.

[Baroni et al., 2018] Baroni, P., Caminada, M., and Giacomin, M. (2018). Abstract argumentation frameworks and their semantics. In Baroni, P., Gabbay, D., Giacomin, M., and van der Torre, L., editors, *Handbook of Formal Argumentation*, pages 159–236. College Publications.

[Baroni and Giacomin, 2007] Baroni, P. and Giacomin, M. (2007). On principle-based evaluation of extension-based argumentation semantics. In *Artificial Intelligence*, volume 171, pages 675–700. Elsevier.

[Baroni et al., 2005] Baroni, P., Giacomin, M., and Guida, G. (2005). SCC-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168(1–2):162–210.

[Baumann et al., 2020a] Baumann, R., Brewka, G., and Ulbricht, M. (2020a). Comparing weak admissibility semantics to their Dung-style counterparts–reduct, modularization, and strong equivalence in abstract argumentation. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*, pages 79–88.

[Baumann et al., 2020b] Baumann, R., Brewka, G., and Ulbricht, M. (2020b). Revisiting the foundations of abstract argumentation–semantics based on weak admissibility and weak defense. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2742–2749.

[Baumann and Ulbricht, 2021] Baumann, R. and Ulbricht, M. (2021). Choices and their consequences-explaining acceptable sets in abstract argumentation frameworks. In *KR*, pages 110–119.

[Bengel and Thimm, 2022] Bengel, L. and Thimm, M. (2022). Serialisable semantics for abstract argumentation. In *Proceedings of COMMA 2022*. to appear.

[Caminada, 2015] Caminada, M. (2015). A discussion game for grounded semantics. In *International Workshop on Theory and Applications of Formal Argumentation*, pages 59–73. Springer.

[Caminada, 2018] Caminada, M. (2018). Argumentation semantics as formal discussion. In Baroni, P., Gabbay, D., Giacomin, M., and van der Torre, L., editors, *Handbook of formal argumentation*, pages 487–518. College Publications.

[Čyras et al., 2021] Čyras, K., Rago, A., Albini, E., Baroni, P., and Toni, F. (2021). Argumentative xai: a survey. *arXiv preprint arXiv:2105.11266*.

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

[Fan and Toni, 2015] Fan, X. and Toni, F. (2015). On computing explanations in argumentation. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

[Thimm, 2022] Thimm, M. (2022). Revisiting initial sets in abstract argumentation. *Argument & Computation*.

[Ulbricht and Wallner, 2021] Ulbricht, M. and Wallner, J. P. (2021). Strong explanations in abstract argumentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 6496–6504.

[Xu and Cayrol, 2016] Xu, Y. and Cayrol, C. (2016). Initial sets in abstract argumentation frameworks. In *Proceedings of the 1st Chinese Conference on Logic and Argumentation (CLAR'16)*, volume 1811, pages 72–85.