Doing Argumentation Theory in Modal Logic [*]

Davide Grossi

Institute of Logic, Language and Computation

d.grossi@uva.nl

Abstract

The present paper applies well-investigated modal logics to provide formal foundations to specific fragments of argumentation theory. This logic-driven analysis of argumentation allows: first, to systematize several results of argumentation theory reformulating them within suitable formal languages; second, to import several techniques (calculi, model-checking, evaluation games, bisimulation games); third, to import results (eminently completeness of axiomatizations, and complexity of model-checking) from modal logic to argumentation theory.

1 Introduction

The present paper analyzes argumentation theory by means of logical tools developed in modal logic. It shows how standard results in argumentation theory obtain elegant reformulations within well-investigated modal logics. This allows to import a number of techniques (e.g., calculi, logical games) as well as results (e.g. completeness, complexity) from modal logic to argumentation theory, essentially for free. Also, as it is often the case in the cross-fertilization of different formalism, such perspective opens up interesting lines of research which were thus far hidden to the attention of argumentation theorists. As such, the present study can be regarded as a study in logic applied to the formal foundations of argumentation theory.

Let us start off with the basic notion of argumentation theory. An abstract argumentation framework is a relational structure $\mathcal{A} = (A, \rightarrow)$ where A is a non-empty set, and $\rightarrow \subseteq A^2$ is a relation on A [9]. This paper investigates the simple but yet unexplored idea which consists in viewing Dung's abstract argumentation frameworks as Kripke frames (W, R) [1]. Modal languages are logical languages which are particularly suitable for talking about relational structures [2] so, from the point of view of this paper, Dung's argumentation frameworks are nothing but Kripke relational frames where the set of arguments A is the

[*] `modalarg.tex`, *Thursday 11th June, 2009, 17:38.*

set of modal states W, and the attack relation $\rightarrow$ is the accessibility relation R. The entire content of the paper hinges on this simple observation.

The paper is structured as follows. Section 2 introduces a well-known modal logic—logic K with converse relation—as a logic for talking about argumentation frameworks. Section 3 uses this logic to formalize a first set of argumentation-theoretic notions such as acceptability, complete and stable extensions. The exposition of such notion will as much as possible stick to [9], in order to emphasize the easiness of modal languages in capturing the natural intuitions backing argumentation theory. As we will see, however, the formalization of such notions can be done only in the meta-language. Section 4 moves on by introducing the further expressivity needed to express argumentation theory in the object language. This enables the possibility of using calculi to derive argumentation-theoretic results such as the Fundamental Lemma [9], and import complexity results concerning, for instance, checking whether an argument belongs to the stable extension of a framework under a given labeling. Along the same line, Section 5 tackles the formalization of the notion of grounded extension within μ-calculus. In Section 6 semantic games are studied for the logic introduced in Section 4 which provide a systematization of dialogue games as model-checking games. Finally, Section 7 tackles the question—not yet addressed in the literature on argumentation theory—of when two arguments, or two argumentation frameworks, are “the same”. In order to shed light on this question the model-theoretic notion of bisimulation is deployed and bisimulation games are introduced as a procedural method to check the “behavioral equivalence” of two argumentation frameworks. Conclusions follow in Section 8 where future research lines are also sketched. Appendix B recapitulates the basic notions of argumentation theory dealt with in the paper.

2 A modal toolkit for argumentation

This section introduces the modal view of argumentation theory investigated in the paper.

2.1 Argumentation models

Doing argumentation theory *à la Dung* means, essentially, to study specific properties of sets of arguments (e.g., conflict-freeness, acceptability, etc.) within a given argumentation framework $\mathcal{A}$. Once an argumentation framework is viewed as a Kripke frame we can directly import the simple machinery deployed by Modal Logic to talk about sets, that is, valuation functions. Modal languages talk about sets by assigning them names, i.e., the atoms of a propositional language, and then by inductively extending such assignments.

Definition 1 (Argumentation models)**.** *Let* $\mathbf{P}$ *be a set of propositional atoms. An argumentation model* $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ *is a structure such that:*

- $\mathcal{A} = (A, \rightarrow)$ *is an argumentation framework;*
- $\mathcal{I} : \mathbf{P} \longrightarrow 2^A$ *is an assignment from* $\mathbf{P}$ *to subsets of* A.

The set of all argumentation models is called $\mathfrak{A}$. *A pointed argumentation model is a pair* $(\mathcal{M}, a)$ *where* $\mathcal{M}$ *is an argumentation model and* a *an argument.*

Argumentation models are nothing but argumentation frames together with a way of "naming" sets of arguments or, to put it otherwise, of "labeling" arguments. In other words, they make explicit the language which is used for talking about sets of arguments. The fact that an argument a belongs to $\mathcal{I}(p)$ in a given model $\mathcal{M}$, which in logical notation reads:

$$(\mathcal{A}, \mathcal{I}), a \models p \tag{1}$$

can be interpreted as stating that "argument a has property p" , or that "p is true of a".

By substituting atom p in Formula 1 with a Boolean compound φ (i.e., $\varphi := p \wedge q$) we can say that "a belongs to both the sets called p and q", and the same can be done for all other Boolean connectives. However, what is typically interesting in argumentation theory, are statements of the sort: "argument a attacks (or is attacked by) an argument in a set called φ". These are modal statements, and the next section introduces the formal language needed for expressing them.

Example 1. *(Argument labelings as argumentation models)* If argumentation frameworks can be viewed as Kripke frames, then an argumentation framework together with a labelling function—in the sense of [4]—from the set $\{\mathtt{1}, \mathtt{0}, ?\}$ is nothing but an argumentation model $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ (Definition 1) where $\mathcal{I}$ interprets the alphabet $\{\mathtt{1}, \mathtt{0}, ?\}$ on the set of arguments A. That is:

- $\mathcal{A} = (A, \rightarrow)$ is an argumentation framework;
- $\mathcal{I}$ is a valuation function from the set of atoms $\mathbf{P} = \{\mathtt{1}, \mathtt{0}, ?\}$ to the set 2^A;
- $\mathcal{M} \models \mathtt{Fct}$, where $\mathtt{Fct} := (\mathtt{1} \wedge \neg\mathtt{0} \wedge \neg ?) \vee (\neg\mathtt{1} \wedge \mathtt{0} \wedge \neg ?) \vee (\neg\mathtt{1} \wedge \neg\mathtt{0} \wedge ?)$. That is, $\mathcal{I}$ is forced to simulate a *function* from A to $\{\mathtt{1}, \mathtt{0}, ?\}$.

We will come back later to the sort of labeling used in argumentation theory to characterize extensions, and show that they can be expressed by modal formulae.

2.2 A basic modal logic for argumentation

We here introduce a first stadard modal logic for talking about the sort of structures introduced in Definition 1.

2.2.1 Language.

Let us now formally introduce the modal language we are going to work with, which we call $\mathcal{L}^{\mathsf{K}^{-1}}$. It consists of a countable set $\mathbf{P}$ of propositional atoms, the set of Boolean connectives $\{\bot, \neg, \wedge\}$, and the set of modal operators $\{\langle\rightarrow\rangle, \langle\leftarrow\rangle\}$. The set of well-formed formulae φ is defined by the following BNF:

$$\mathcal{L}^{\mathsf{K}^{-1}} : \varphi ::= p \mid \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\rightarrow\rangle\varphi \mid \langle\leftarrow\rangle\varphi$$

where p ranges over $\mathbf{P}$. The other standard boolean connectives $\{\top, \vee, \rightarrow\}$, and the modal duals $\{[\rightarrow], [\leftarrow]\}$ are defined as usual.

We can now express that "a attacks an argument belonging to a set called φ" (Formula 2), that "a is attacked by an argument in a set called φ" (Formula

3), or that "*a reinstates* an argument in φ" (Formula 3) in the sense that it attacks an attacker of a φ argument, or that "*a* is *defended* by the set φ" (Formula 3):

$$
\begin{align}
(\mathcal{A}, \mathcal{I}), a &\models \langle \rightarrow \rangle \varphi \tag{2}\\
(\mathcal{A}, \mathcal{I}), a &\models \langle \leftarrow \rangle \varphi \tag{3}\\
(\mathcal{A}, \mathcal{I}), a &\models \langle \rightarrow \rangle \langle \rightarrow \rangle \varphi \tag{4}
\end{align}
$$

The next section makes these intuitive readings exact by defining the formal semantics of $\mathcal{L}^{\mathsf{K}^{-1}}$ in terms of argumentation models.

2.2.2 Semantics.

The formal semantics of $\mathcal{L}^{\mathsf{K}^{-1}}$ is defined as usual via the notion of satisfaction of a formula in a model.

Definition 2 (Satisfaction for $\mathcal{L}^{\mathsf{K}^{-1}}$ in argumentation models)**.** *Let $\varphi \in \mathcal{L}^{\mathsf{K}^{-1}}$. The satisfaction of φ by a pointed argumentation model $(\mathcal{M}, a)$ is inductively defined as follows:*

$$
\begin{array}{rll}
\mathcal{M}, a \not\models \bot & & \\
\mathcal{M}, a \models p & \textit{iff} & a \in \mathcal{I}(p), \textit{ for } p \in \mathbf{P}\\
\mathcal{M}, a \models \neg\varphi & \textit{iff} & \mathcal{M}, a \not\models \varphi\\
\mathcal{M}, a \models \varphi_1 \wedge \varphi_2 & \textit{iff} & \mathcal{M}, a \models \varphi_1 \text{ AND } \mathcal{M}, a \models \varphi_2\\
\mathcal{M}, a \models \langle \rightarrow \rangle \varphi & \textit{iff} & \exists b \in A : (a, b) \in \rightarrow \text{ AND } \mathcal{M}, b \models \varphi\\
\mathcal{M}, a \models \langle \leftarrow \rangle \varphi & \textit{iff} & \exists b \in A : (a, b) \in \rightarrow^{-1} \text{ AND } \mathcal{M}, b \models \varphi
\end{array}
$$

As usual, the truth-set of φ in model $\mathcal{M}$ is denoted $||\varphi||_{\mathcal{M}}$.[1] *We say that: φ is valid in an argumentation model $\mathcal{M}$ iff it is satisfied in all pointed models of $\mathcal{M}$, i.e., $\mathcal{M} \models \varphi$; φ is valid in a class $\mathfrak{M}$ of argumentation models iff it is valid in all its models, i.e., $\mathfrak{M} \models \varphi$. All definitions are naturally generalizable to sets of formulae Φ.*

Let us comment upon the two modal clauses. A formula $\langle \rightarrow \rangle \varphi$ is satisfied by argument a in model $\mathcal{M}$ if and only if there exists an argument b such that *a attacks b* and b belongs to the set $||\varphi||_{\mathcal{M}}$. Conversely, a formula $\langle \leftarrow \rangle \varphi$ is satisfied by argument a in model $\mathcal{M}$ if and only if there exists an argument b such that *a is attacked* by b and b belongs to the set $||\varphi||_{\mathcal{M}}$. In other words $\langle \leftarrow \rangle$ is interpreted on the inverse $\rightarrow^{-1}$ of the attack relation $\rightarrow$.

Definition 2 provides a structured way to define sets of arguments by means of expressions of $\mathcal{L}^{\mathsf{K}^{-1}}$. If an argument belongs to a set specified by φ in $\mathcal{M}$, that is $a \in ||\varphi||_{\mathcal{M}}$, then we write $\mathcal{M}, a \models \varphi$ and we say that a satisfies φ or that a is a φ-argument.

The set of formulae φ of $\mathcal{L}^{\mathsf{K}^{-1}}$ such that $\mathfrak{A} \models \varphi$, defines logic K^{-1}. Such logic contains all the truths concerning argumentation frameworks which can be expressed in $\mathcal{L}^{\mathsf{K}^{-1}}$. The next section introduces a Hilbert calculus for this logic.

[1] Subscript $\mathcal{M}$ will often be dropped when no confusion arises.

2.2.3 Axiomatics.

Logic K^{-1} is axiomatized by the following set of schemata and rules:

`(Prop)`	propositional schemata
`(K)`	$[i](\varphi_1 \rightarrow \varphi_2) \rightarrow ([i]\varphi_1 \rightarrow [i]\varphi_2)$
`(Conv)`	$\varphi \rightarrow [i]\neg[j]\neg\varphi$
`(Dual)`	$\langle i \rangle \leftrightarrow \neg[i]\neg\varphi$
`(MP)`	IF $\vdash \varphi_1 \rightarrow \varphi_2$ AND $\vdash \varphi_1$ THEN φ_2
`(N)`	IF $\vdash \varphi$ THEN $\vdash [i]\varphi$

with $i \neq j \in \{\rightarrow, \leftarrow\}$. We have the following result.

2.2.4 Meta-theoretical results.

We have the following results:

- Logic K^{-1} is sound and strongly complete with respect to the class $\mathfrak{A}$ of all argumentation models under the semantics given in Definition 2 (see Appendix for a sketch of the proof).
- The satisfiability problem of K^{-1} is P-reducible to the one of K in the presence of a background theory [11], which is known to be EXP-complete [19].

In the next section the logic just introduced is used to start off with a formalization of some basic argumentation-theoretic notions.

3 Doing argumentation in K^{-1}: basic notions

How much of abstract argumentation can be done within K^{-1}? The present section answers this question. Surprisingly, almost all the key notions introduced by Dung in [9] can be expressed and study resorting to this a simple logic, although only at the level of the meta-language.

3.1 Acceptability, conflict-freeness and admissibility

Given an argumentation model $\mathcal{M}$, an argument is said to be *acceptable with respect to a set* $||\varphi||$ in $\mathcal{M}$ if and only if for all arguments b attacking a there exists one φ-argument c s.t. c attacks b. That is:

$$\mathcal{M}, a \models [\leftarrow]\langle\leftarrow\rangle\varphi \tag{5}$$

In other words, formula $[\leftarrow]\langle\leftarrow\rangle\varphi$ states that for any attack on a there exists a reinstatement from a $||\varphi||$-argument.

Similarly, we can express that a set of arguments $||\varphi||$ is acceptable with respect to a set of arguments $||\psi||$ in model $\mathcal{M}$. This holds if and only if all arguments a in $||\varphi||$ are acceptable with respect to $||\psi||$. That is to say, $||\varphi|| \subseteq ||[\leftarrow]\langle\leftarrow\rangle\psi||$, which in modal logic corresponds to the statement of the following global property:

$$\mathcal{M} \models \varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\psi \tag{6}$$

To put it otherwise, formula $\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\psi$ states that the set of arguments $||\varphi||$ is able to defend all its members from the attack of other arguments (which are also possibly in $||\varphi||$). The notion of self-acceptability is therefore straightforwardly defined:

$$\mathcal{M} \models \varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi \tag{7}$$

Global properties of models such as Formulae 6 and 7 are typical example of the type of notions playing a central role in argumentation theory.

Other global properties of argumentation models which play a key role in Dung's theory are conflict-freeness and admissibility. A set of arguments $||\varphi||$ is said to be *conflict free* in $\mathcal{M}$ iff no argument in $||\varphi||$ attacks any argument in $||\varphi||$:

$$\mathcal{M} \models \varphi \rightarrow \neg\langle\rightarrow\rangle\varphi \tag{8}$$

That is to say, $||\varphi||$ is conflict-free if and only if either an argument does not satisfy φ or, if it is a φ-argument, then it does not attack any φ-argument. It is a matter of direct application of the semantics to prove the following fact.

Fact 1 (Equivalence of → and ← for conflict-freeness)**.** *Let $\mathcal{M}$ be an argumentation model. It holds that:*

$$\mathcal{M} \models \varphi \rightarrow \neg\langle\rightarrow\rangle\varphi \quad \Longleftrightarrow \quad \mathcal{M} \models \varphi \rightarrow \neg\langle\leftarrow\rangle\varphi$$

Proof. [Left to right] We proceed per absurdum.Take $\mathcal{M} \models \varphi \rightarrow \neg\langle\rightarrow\rangle\varphi$ and suppose $\mathcal{M} \not\models \varphi \rightarrow \neg\langle\leftarrow\rangle\varphi$. It follows that there exist arguments a and b such that $b \leftarrow a$ and $\mathcal{M}, a \models \varphi$. However, from the assumption we have that if $\mathcal{M}, a \models \varphi$, then for all arguments b such that $a \rightarrow b$, $\mathcal{M}, b \models \neg\varphi$. We thus obtain a contradiction. [Right to left] An analogous argument per absurdum can be used. □

So, as we might expect, conflict-freeness can be equivalently described either by thinking in terms of arguments attacking other arguments, or by thinking in terms of arguments being attacked by other arguments.

Acceptability and conflict-freeness together determine the *admissibility* of a set of arguments. A set $||\varphi||$ is admissible in $\mathcal{M}$ if and only if it is acceptable in $\mathcal{M}$ with respect to itself, that is, if and only if the following validity holds:

$$\mathcal{M} \models (\varphi \rightarrow \neg\langle\rightarrow\rangle\varphi) \wedge (\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \tag{9}$$

which, by propositional logic, is equivalent to the following slicker formulation:

$$\mathcal{M} \models \varphi \rightarrow ([\rightarrow]\neg\varphi \wedge [\leftarrow]\langle\leftarrow\rangle\varphi) \tag{10}$$

Formulae 9 and 10 state that the set of φ-arguments is such that all its arguments attack arguments that do not belong to $||\varphi||$, and all arguments attacking its arguments are reinstated by other φ-arguments. If this holds for a φ in , in an argumentation model $\mathcal{M}$, then $||\varphi||$ is admissible in $\mathcal{M}$.

Table 1 recapitulates the formalization in K^{-1} of self-acceptability, conflict-freenes and admissibility. All such notions can be captured as validities of $\mathcal{L}^{\mathsf{K}^{-1}}$ formulae in the argumentation model at issue.

$Acc(\varphi, \psi, \mathcal{M})$	$\Longleftrightarrow$	$\mathcal{M} \models \varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\psi$
$CFree(\varphi, \mathcal{M})$	$\Longleftrightarrow$	$\mathcal{M} \models \varphi \rightarrow \neg\langle\rightarrow\rangle\varphi$
$Adm(\varphi, \mathcal{M})$	$\Longleftrightarrow$	$\mathcal{M} \models \varphi \rightarrow ([\rightarrow]\neg\varphi \wedge [\leftarrow]\langle\leftarrow\rangle\varphi)$

Table 1: Acceptability, conflict-freeness and admissibility in $\mathcal{L}^{\mathsf{K}^{-1}}$

3.2 Complete and stable extensions

In [9], the "solution" of an argumentation framework is a set of arguments which can be considered as a "rational position" to be held according to some kind of precisely defined notion of rationality. Two of such solution concepts are the so-called *complete* and *stable* extensions.

Given an argumentation model $\mathcal{M}$, a complete extension of $\mathcal{M}$ is a set $||\varphi||$ which is admissible in $\mathcal{M}$ and is such that any argument which is acceptable for $||\varphi||$ in $\mathcal{M}$ belongs to $||\varphi||$. In $\mathcal{L}^{\mathsf{K}^{-1}}$ this becomes:

$$\mathcal{M} \quad \models \quad \varphi \rightarrow ([\rightarrow]\neg\varphi \wedge [\leftarrow]\langle\leftarrow\rangle\varphi) \wedge ([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow \varphi) \tag{11}$$

which, by propositional logic, is equivalent to:

$$\mathcal{M} \quad \models \quad (\varphi \rightarrow [\rightarrow]\neg\varphi) \wedge (\varphi \leftrightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \tag{12}$$

So, a set of φ-arguments is a complete extension of an argumentation model $\mathcal{M}$ iff such set is conflict-free in $\mathcal{M}$ (first conjunct of Formula 12) and it is equivalent to the set of arguments it defends (second conjunct of Formula 12).

We can similarly capture the notion of stable extension for a given argumentation model $\mathcal{M}$. According to Dung, $||\varphi||$ is a stable extension if and only if $||\varphi||$ is the set of arguments which is not attacked by $||\varphi||$, that is:

$$\mathcal{M} \models \varphi \leftrightarrow \neg\langle\leftarrow\rangle\varphi \tag{13}$$

Table 2 recapitulates the semantic definitions of completeness and stability in K^{-1}. The following fact can be proven by model-theoretic considerations.

Fact 2 (Stability implies admissibility)**.** *Let $\mathcal{M} = (A, \mathcal{I})$ be an argumentation model. It holds that:*

$$Stable(\varphi, \mathcal{M}) \Longrightarrow Adm(\varphi, \mathcal{M}).$$

$Complete(\varphi, \mathcal{M})$	$\Longleftrightarrow$	$\mathcal{M} \models (\varphi \rightarrow [\rightarrow]\neg\varphi) \wedge (\varphi \leftrightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$
$Stable(\varphi, \mathcal{M})$	$\Longleftrightarrow$	$\mathcal{M} \models \varphi \leftrightarrow \neg\langle\leftarrow\rangle\varphi$

Table 2: Complete and stable extensions in $\mathcal{L}^{\mathsf{K}^{-1}}$

Proof. [$Stable(\varphi, \mathcal{M}) \Longrightarrow CFree(\varphi, \mathcal{M})$] We proceed per absurdum. Consider $\mathcal{M} \models \varphi \leftrightarrow \neg\langle\leftarrow\rangle\varphi$ and suppose there exists $a \in A$ such that $\mathcal{M}, a \models \varphi \wedge \langle\rightarrow\rangle\varphi$. Then there exists $b \in A$ such that $a \rightarrow b$ and $\mathcal{M}, b \models \varphi$, which is impossible since $\mathcal{M}, b \models \neg\langle\leftarrow\rangle\varphi$ by assumption. [$Stable(\varphi, \mathcal{M}) \Longrightarrow Acc(\varphi, \varphi, \mathcal{M})$] We proceed again per absurdum. Consider the contrapositive of Formula 13, i.e., $\mathcal{M} \models \neg\varphi \leftrightarrow \langle\leftarrow\rangle\varphi$, and suppose there exists $a \in A$ such that $\mathcal{M}, a \models \varphi \wedge \neg[\leftarrow]\langle\leftarrow\rangle\varphi$. It follows that there exists a$b \in A$ such that $a \leftarrow b$ and $\mathcal{M}, b \models \neg\varphi \wedge [\leftarrow]\neg\varphi$. From this, by our assumption, it follows that $\mathcal{M}, b \models \langle\leftarrow\rangle\varphi \wedge [\leftarrow]\neg\varphi$, which is impossible. □

Fact 2 shows how model-theoretic properties of K^{-1} reflect basic theorems of abstract argumentation. It is worth noticing that the proof of this fact cannot be carried out as a derivation within K^{-1} since it lacks the necessary expressivity to represent validity within a model as a formula in the object language (e.g., the universal modality [1]). A more expressive logic where this can be done is exposed in Appendix. Here we have opted for a simpler formalism which can better illustrate the methodology behind our work.

3.3 Characteristic functions and K^{-1}

Each argumentation framework $\mathcal{A} = (A, \rightarrow)$ determines a *characteristic function* $c_{\mathcal{A}} : 2^A \longrightarrow 2^A$ such that for any set of arguments X, $c_{\mathcal{A}}(X)$ yields the set of arguments in A which are acceptable with respect to X, i.e., $\{a \in A \mid \forall b \in A : [b \rightarrow a \Rightarrow \exists c \in X : c \rightarrow b]\}$.

Now, consider language $\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}$ defined by the following BNF:

$$\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle} : \varphi ::= p \mid \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid [\leftarrow]\langle\leftarrow\rangle\varphi$$

where p belongs to the set of atoms $\mathbf{P}$. Notice that $\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}$ is the fragment of $\mathcal{L}^{\mathsf{K}^{-1}}$ containing only the compounded modal operator $[\leftarrow]\langle\leftarrow\rangle$. Let $\mathcal{A}^+ = (2^A, \cap, -, \emptyset, c_{\mathcal{A}})$ be the power set algebra on 2^A extended with operator $c_{\mathcal{A}}$, and consider the term algebra $\mathfrak{ter}_{\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}} = (\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}, \wedge, \neg, \bot, [\leftarrow]\langle\leftarrow\rangle)$. We can prove the following interesting fact.

Theorem 1 ($c_{\mathcal{A}}$ vs. $[\leftarrow]\langle\leftarrow\rangle$)**.** *Let $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ be an argumentation model. The restriction $\mathcal{I}\lceil\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}$ of the interpretation function $\mathcal{I}$ is a homomorphism from $\mathfrak{ter}_{\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}}$ to $\mathcal{A}^+$.*

Proof. The case of Boolean connectives is trivial. It remains to be proven that for any φ: $||[\leftarrow]\langle\leftarrow\rangle\varphi||_{\mathcal{M}} = c_{\mathcal{A}}(||\varphi||_{\mathcal{M}})$. It suffices to spell out the semantics of $[\leftarrow]\langle\leftarrow\rangle$ recalling that $\leftarrow = \rightarrow^{-1}$:

$$\begin{aligned} ||[\leftarrow]\langle\leftarrow\rangle\varphi||_{\mathcal{M}} &= \{a \in A \mid \forall b : a \leftarrow b, \exists c : b \leftarrow c \text{ and } c \in ||\varphi||_{\mathcal{M}}\} \\ &= \{a \in A \mid \forall b : b \rightarrow a, \exists c : c \rightarrow b \text{ and } c \in ||\varphi||_{\mathcal{M}}\} \\ &= c_{\mathcal{A}}(||\varphi||_{\mathcal{M}}). \end{aligned}$$

This completes the proof. □

In other words, Fact 1 shows that the complex modal operator $[\leftarrow]\langle\leftarrow\rangle$, under the semantics provided in Definition 2, behaves exactly like the characteristic function of the argumentation frameworks on which the argumentation models are built. To put it yet otherwise, formulae of the form $[\leftarrow]\langle\leftarrow\rangle\varphi$ denote the value of the characteristic function applied to the set of φ-arguments.

From Theorem 1 it becomes thus clear that: a self-acceptable set of arguments $||\varphi||$ is a set for which $[\leftarrow]\langle\leftarrow\rangle$ increases, i.e., $||\varphi|| \subseteq ||[\leftarrow]\langle\leftarrow\rangle\varphi||$ (Formula 5); an admissible set of arguments $||\varphi||$ is a conflict-free set for which $[\leftarrow]\langle\leftarrow\rangle$ is increasing (Formula 9); a complete extension $||\varphi||$ is a fixpoint of $[\leftarrow]\langle\leftarrow\rangle$, i.e., $||\varphi|| = ||[\leftarrow]\langle\leftarrow\rangle\varphi||$ (Formula 11). All such statements are counterparts of statements to be found in [9]. We can now study the properties of $[\leftarrow]\langle\leftarrow\rangle\varphi$ by resorting to the semantics of K^{-1}.

Fact 3 (Model-theoretic properties of $[\leftarrow]\langle\leftarrow\rangle$)**.** *Let $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ be an argumentation model and $\mathcal{M}^s = (\mathcal{A}^s, \mathcal{I})$ a serial argumentation model, that is, such that $\rightarrow^{-1}$ in $\mathcal{A}^s$ is serial. It holds that:*

$$\textbf{Monotonicity}: \quad \mathcal{M} \models \varphi_1 \rightarrow \varphi_2 \Longrightarrow \mathcal{M} \models [\leftarrow]\langle\leftarrow\rangle\varphi_1 \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi_2$$

$$\textbf{Normality}: \quad \mathcal{M}^s \models \varphi \rightarrow \bot \Longrightarrow \mathcal{M}^s \models [\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow \bot$$

Proof. [Monotonicity] Let us proceed per absurdum, assuming that $\mathcal{M} \models \varphi_1 \rightarrow \varphi_2$ and $\mathcal{M} \not\models [\leftarrow]\langle\leftarrow\rangle\varphi_1 \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi_2$. This latter means that there exists $a \in A$ such that $\mathcal{M}, a \models [\leftarrow]\langle\leftarrow\rangle\varphi_1 \wedge \langle\leftarrow\rangle[\leftarrow]\neg\varphi_2$ which in turn implies the existence of $b \in A$ such that $\mathcal{M}, b \models \langle\leftarrow\rangle\varphi_1 \wedge [\leftarrow]\neg\varphi_2$. Given the assumption this is impossible. [Normality] It can be proven directly. Assume $\mathcal{M}^s \models \varphi \rightarrow \bot$ and $\mathcal{M}^s \models [\leftarrow]\langle\leftarrow\rangle\varphi$. It follows that $\mathcal{M}^s \models [\leftarrow]\langle\leftarrow\rangle\bot$ which is impossible since $\rightarrow^{-1}$ is serial in $\mathcal{M}^s$. Hence $\mathcal{M}^s \models [\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow \bot$. □

Monotonicity guarantees that the set of arguments reinstating arguments in a given set $||\varphi||$ grows if $||\varphi||$ grows. Normality states that in a serial argumentation model the set of arguments which is acceptable with respect to the empty set, i.e., $||\bot||$, is empty.[2]

4 Argumentation in K^{U}: universal modality

The previous section has introduced a modal logic for talking about the relations of "attacking" and "being attacked by". However, as shown in Table 1 and 2, and on the ground of Fact 1, the only relation occurring in the formalization of the argumentation theoretic notions considered is the relation $\leftarrow$, i.e., "being attacked by". In this section, we restrict K^{-1} to its "being attacked by" fragment—thus allowing only the $\langle\leftarrow\rangle$ and $[\leftarrow]$ modal operators—and extend it with the universal modality [1]. The resulting system is nothing but K^{U}, that is, the minimal normal modal logic K extended with the universal modality.

4.1 Logic K^{U}

Logic K^{U} is a well-investigated system. In this section we recapitulate its semantics, axiomatics and some of its meta-logical properties.

[2] It might be instructive to notice that seriality implies non well-foundedness since if $\rightarrow^{-1}$ is serial, every argument has a $\rightarrow^{-1}$-successor.

4.1.1 Language.

As anticipated above, the language of K^{U} is a standard modal language built on the set of atoms $\mathbf{P}$ by the following BNF:

$$\mathcal{L}^{\mathsf{K}^{\mathsf{U}}} : \varphi ::= p \mid \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\leftarrow\rangle\varphi \mid \langle\mathsf{U}\rangle\varphi$$

where p ranges over $\mathbf{P}$. The other standard boolean connectives $\{\top, \vee, \rightarrow\}$, and the modal duals $\{[\leftarrow], [\mathsf{U}]\}$ are defined as usual.

Logic K^{U} is therefore endowed with modal operators of the type "there exists an argument attacking the current one such that"—$\langle\leftarrow\rangle$—and "there exists an argument such that"—$\langle\mathsf{U}\rangle$—together with their duals.

4.1.2 Semantics.

The semantics of K^{U} extends the one of K^{-1} (Definition 2) with the clause for the universal modality.

Definition 3 (Satisfaction for $\mathcal{L}^{\mathsf{K}^{\mathsf{U}}}$ in argumentation models)**.** *Let $\varphi \in \mathcal{L}^{\mathsf{K}^{\mathsf{U}}}$. The satisfaction of φ by a pointed argumentation model $(\mathcal{M}, a)$ is inductively defined as follows (Boolean clauses are omitted):*

$$\begin{array}{lll} \mathcal{M}, a \models \langle\leftarrow\rangle\varphi & \textit{iff} & \exists b \in A : (a, b) \in \rightarrow^{-1} \text{ AND } \mathcal{M}, b \models \varphi \\ \mathcal{M}, a \models \langle\mathsf{U}\rangle\varphi & \textit{iff} & \exists b \in A : \mathcal{M}, b \models \varphi \end{array}$$

We say that: φ is valid in an argumentation model $\mathcal{M}$ iff it is satisfied in all pointed models of $\mathcal{M}$, i.e., $\mathcal{M} \models \varphi$; φ is valid in a class $\mathfrak{M}$ of argumentation models iff it is valid in all its models, i.e., $\mathfrak{M} \models \varphi$. All definitions are naturally generalizable to sets of formulae Φ.

In words, what K^{U} adds to K^{-1} is existential and universal quantification via the universal modalities $\langle\mathsf{U}\rangle$ and $[\mathsf{U}]$.

4.1.3 Axiomatics.

The logic K^{U} is axiomatized as follows:

(**Prop**)	propositional tautologies
(**K**)	$[i](\varphi_1 \rightarrow \varphi_2) \rightarrow ([i]\varphi_1 \rightarrow [i]\varphi_2)$
(**T**)	$[\mathsf{U}]\varphi \rightarrow \varphi$
(**4**)	$[\mathsf{U}]\varphi \rightarrow [\mathsf{U}][\mathsf{U}]\varphi$
(**5**)	$\neg[\mathsf{U}]\varphi \rightarrow [\mathsf{U}]\neg[\mathsf{U}]\varphi$
(**Incl**)	$[\mathsf{U}]\varphi \rightarrow [i]\varphi$
(**Dual**)	$\langle i\rangle\varphi \leftrightarrow \neg[i]\neg\varphi$

with $i \in \{\leftarrow, \mathsf{U}\}$.

4.1.4 Meta-theoretical results.

We list the following known results, which are relevant for our purposes.

- Logic K^{U} is sound and strongly complete for the class $\mathfrak{A}$ of argumentation frames [1, Ch. 7].
- The complexity of deciding whether a formula of $\mathcal{L}^{\mathsf{K}^{\mathsf{U}}}$ is satisfiable is EXP-complete [14].
- The complexity of checking whether a formula of $\mathcal{L}^{\mathsf{K}^{\mathsf{U}}}$ is satisfied by a pointed model $\mathcal{M}$ is P-complete [13].

4.2 Doing argumentation in K^{U}

We have now a calculus which fits very well with argumentation models. The present section shows how such calculus, and its semantics, can be concretely deployed to express basic notion of argumentation theory in a formal language, and consequently obtain formal proofs of theorems of argumentation theory.

Logic K^{U} is expressive enough to capture the following notions in the object-language.

$$\begin{aligned}
Acc(\varphi,\psi) &:= [\mathsf{U}](\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\psi) && (14)\\
CFree(\varphi) &:= [\mathsf{U}](\varphi \rightarrow \neg\langle\leftarrow\rangle\varphi) && (15)\\
Adm(\varphi) &:= [\mathsf{U}](\varphi \rightarrow ([\leftarrow]\neg\varphi \wedge [\leftarrow]\langle\leftarrow\rangle\varphi)) && (16)\\
Complete(\varphi) &:= [\mathsf{U}]((\varphi \rightarrow [\leftarrow]\neg\varphi) \wedge (\varphi \leftrightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)) && (17)\\
Stable(\varphi) &:= [\mathsf{U}](\varphi \leftrightarrow \neg\langle\leftarrow\rangle\varphi) && (18)
\end{aligned}$$

Notice that these definitions restate the meta-language definitions summarized in Tables 1 and 2.

Example 2. *(Argumentation labelings in K^{U})* According to [4], an argumentation labeling $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ is a *complete labeling* if and only if for each $a \in A$:

1. $\mathcal{M}, a \models 1$ if and only if for all b s.t. $a \leftarrow b$, $\mathcal{M}, b \models \mathbb{0}$;
2. $\mathcal{M}, a \models \mathbb{0}$ if and only if there exists b s.t. $a \leftarrow b$ and $\mathcal{M}, b \models 1$
3. $\mathcal{M} \models \texttt{Fct}$ (see Example 1).

It is striking how such conditions—in particular 1 and 2—exhibit a natural modal flavor. Here it is their reformulation in K^{U}:

$$Complete(\mathcal{M}) \quad := \quad \mathcal{M} \models [\mathsf{U}]((1 \leftrightarrow [\leftarrow]\mathbb{0}) \wedge (\mathbb{0} \leftrightarrow \langle\leftarrow\rangle 1) \wedge \texttt{Fct}) \qquad (19)$$

We have the following fact.

Fact 4. *Let $\mathcal{M}$ be an argumentation model for $\mathbf{P} = \{1, \mathbb{0}, ?\}$. It holds that:*

$$\begin{aligned}
Complete(\mathcal{M}) \quad \Longleftrightarrow \quad & \mathcal{M} \models Complete(1) \wedge [\mathsf{U}]\texttt{Fct}\\
& \wedge [\mathsf{U}](? \leftrightarrow (\langle\leftarrow\rangle\neg\mathbb{0} \wedge \neg\langle\leftarrow\rangle 1))
\end{aligned}$$

Proof. From left to right. Follows directly from Formula 19. From right to left. It also follows directly from Formula 19 by considering that if $\mathcal{M}, a \models 1$ then $\mathcal{M}, a \models [\leftarrow]\mathbb{0}$ since otherwise $\mathcal{M}, a \models ?$ which is incompatible with the validity of $\texttt{Fct}$. Similarly, if $\mathcal{M}, a \models \langle\leftarrow\rangle 1$ then $\mathcal{M}, a \models \mathbb{0}$ by the validity of $? \leftrightarrow (\langle\leftarrow\rangle\neg\mathbb{0} \wedge \neg\langle\leftarrow\rangle 1)$. □

In other words, the characterization of complete extensions in terms of labellings coincides with the characterization of complete extensions in terms of truth-sets. Notice also that, as a corollary, we obtain that the existence of a complete labeling implies the existence of a complete extension and vice versa, the existence of a model where $||\varphi||$ is a complete extension implies the existence of a complete labeling, since any model can be extended to the vocabulary $\{\mathtt{1}, \mathtt{0}, ?\}$ and constrained in order to satisfy $\mathtt{Fct}$ and $? \leftrightarrow (\langle\leftarrow\rangle\neg\mathtt{0} \wedge \neg\langle\leftarrow\rangle\mathtt{1})$. Similar characterizations, which are a typical asset of the labelling-based approach to argumentation, can be obtained for all the notions formalized in Formulae 14-18.

Logic K^{U} has therefore sufficient expressive power to capture a number of central results of argumentation theory. In this section we provide a sample of such results taken from [9], formalized and proved within K^{U}.

Theorem 2 (Fundamental Lemma)**.** *The following formula is a theorem of* K^{U}*:*

$$Adm(\varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow Adm(\varphi \vee \psi) \wedge Acc(\xi, \varphi \vee \psi) \tag{20}$$

Sketch. The desired validity can be proven syntactically by then resorting to soundness. We provide, as an example, the derivation of a sub-result, namely, $Acc(\varphi, \varphi) \wedge Acc(\psi, \varphi) \rightarrow Acc(\varphi \vee \psi, \varphi \vee \psi)$. Notice that the antecedent and consequent of this implication are implied by the antecedent and, respectively, the consequent of Formula 20.

$$\begin{array}{lll}
1. & ((\alpha \rightarrow \gamma) \wedge (\beta \rightarrow \gamma)) \rightarrow (\alpha \vee \beta \rightarrow \gamma) & \textbf{Prop} \\
2. & ([\mathsf{U}](\alpha \rightarrow \gamma) \wedge [\mathsf{U}](\beta \rightarrow \gamma)) \rightarrow [\mathsf{U}](\alpha \vee \beta \rightarrow \gamma) & 2, \textbf{N}, \textbf{K}, \textbf{MP} \\
3. & ([\mathsf{U}](\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \wedge [\mathsf{U}](\psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)) \rightarrow & \\
 & [\mathsf{U}](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) & \text{Instance of 3} \\
4. & [\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle(\varphi \vee \psi) & \textbf{Prop}, \textbf{K}, \textbf{N} \\
5. & ([\mathsf{U}](\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \wedge [\mathsf{U}](\psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)) \rightarrow & \\
 & [\mathsf{U}](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi \vee \psi) & 4, \textbf{Prop}, \textbf{K}, \textbf{N} \\
6. & Acc(\varphi, \varphi) \wedge Acc(\psi, \varphi) \rightarrow Acc(\varphi \vee \psi, \varphi \vee \psi) & 5, \text{definition}
\end{array}$$

The proof is completed by proving that $Adm(\varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow CF(\varphi \vee \psi)$ and that $Adm(\varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow Acc(\xi, \varphi \vee \psi)$. □

Notice that Theorem 2 is, in fact, a generalized version of the Fundamental Lemma proven in [9]. We provide one more example of theorems of abstract argumentation which can be obtained as formal theorems of K^{U}.

Theorem 3 (Stable implies admissible and complete)**.** *The following formulae are theorems of* K^{U}*:*

$$Stable(\varphi) \rightarrow Adm(\varphi) \tag{21}$$

$$Stable(\varphi) \rightarrow Complete(\varphi) \tag{22}$$

Proof. Formula 21 follows from Fact 2 and the completeness of K^{U}. Formula 22 is a direct corollary of Formula 21, the definition of *Stable*(φ), the definition of *Complete*(φ) and the completeness of K^{U}. □

Other results can be formalized along the same lines. What this section aimed at showing is that, already within a rather standard modal systems such as K^{U}, quite many notions and results of abstract argumentation can be accommodated. The next section shows what kind of modal machinery is needed to capture the notion of *grounded extension* which we have not yet discussed.

5 Argumentation in K^{μ}: least fixpoints

Let us go back for a moment to logic K^{-1}, and to the way its $[\leftarrow]\langle\leftarrow\rangle$-formulae formalizing the notion of characteristic function of a given argumentation model (Section 3.3). Carrying on with the analogy, we have that a formula φ is a $[\leftarrow]\langle\leftarrow\rangle$-fixpoint for an argumentation model $\mathcal{M}$ if and only if $\mathcal{M} \models \varphi \leftrightarrow [\leftarrow]\langle\leftarrow\rangle\varphi$. We have the following.

Corollary 1 (Existence of $[\leftarrow]\langle\leftarrow\rangle$-fixpoints)**.** *For every argumentation model* $\mathcal{M}$*, there exist a greatest and a least* $[\leftarrow]\langle\leftarrow\rangle$*-fixpoint.*

Proof. The result follows from Theorem 1 and Fact 3 via a direct application of the Knaster-Tarski fixpoint theorem[3] on $\mathtt{ter}_{\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}} = (\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}, \wedge, \neg, \bot, [\leftarrow]\langle\leftarrow\rangle)$. □

Logic K^{-1} does not have the necessary expressive power to talk about greatest and least fixpoints for $[\leftarrow]\langle\leftarrow\rangle$. In the next section, we enhance K^{-1} with fixpoint operators, thus moving into the realm of the so-called μ-calculi [3].

5.1 A μ-calculus for argumentation

The present section introduces the μ-calculus in the context of argumentation theory.

5.1.1 Language.

As already noticed at the beginning of Section 4, we can profitably restrict $\mathcal{L}^{\mathsf{K}^{-1}}$ to its "being attacked" part when it comes to expressing traditional notions, that is, operators $\langle\leftarrow\rangle$ and $[\leftarrow]$. We introduce the least fixpoint operator μ on the top of this language, and define language $\mathcal{L}^{\mathsf{K}^{\mu}}$ via the following BNF:

$$\mathcal{L}^{\mathsf{K}^{\mu}} : \varphi ::= p \mid \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\leftarrow\rangle\varphi \mid \mu p.\varphi(p)$$

where p ranges over $\mathbf{P}$ and $\varphi(p)$ indicates that p occurs free in φ (i.e., it is not bounded by fixpoint operators) and under an even number of negations.[4] In general, the notation $\varphi(\psi)$ stands for ψ occurs in φ. The usual definitions

[3] We refer the interested reader to [7] for a neat formulation of this result.

[4] This syntactic restriction guarantees that every formula $\varphi(p)$ defines a set transformation which preserves $\subseteq$, which in turn guarantees the existence of least and greatest fixpoints by the Knaster-Tarski fixpoint theorem.

for Boolean and modal operators can be applied. Also, the greatest fixpoint operator ν can be defined from μ as follows: $\nu x.\varphi(x) := \neg\mu x.\neg\varphi(\neg x)$.

Intuitively, $\mu x.\varphi(x)$ denotes the smallest formula x such that $x \leftrightarrow \varphi(x)$. To immediately appreciate the usefulness of such operator in our context, take $\varphi(x) := [\leftarrow]\langle\leftarrow\rangle x$, that is, take $\varphi(x)$ to be the modal version $[\leftarrow]\langle\leftarrow\rangle$ of the characteristic function, and apply it to formula x. What we obtain is a modal formula expressing the least fixpoint of a characteristic function:

$$\mu p.[\leftarrow]\langle\leftarrow\rangle p \tag{23}$$

Language $\mathcal{L}^{\mathsf{K}^\mu}$ can therefore express the least complete extension as a modal formula. We further investigate the expressivity of this language for argumentation theoretic purposes in Section 5.2. First, however, we spell out the semantics of $\mathcal{L}^{\mathsf{K}^\mu}$, and provide a sound and complete axiomatization of the logic thus obtained.

5.1.2 Semantics.

The semantics of μ-calculi is most perspicuously given in an algebraic fashion, which is what we do in the next definition.

Definition 4 (Satisfaction for $\mathcal{L}^{\mathsf{K}^\mu}$ in argumentation models)**.** *Let $\varphi \in \mathcal{L}^{\mathsf{K}^\mu}$. The satisfaction of φ by a pointed argumentation model $(\mathcal{M}, a)$ is inductively defined as follows:*

$$\begin{array}{rll}
\mathcal{M}, a \not\models \bot & & \\
\mathcal{M}, a \models p & \textit{iff} & a \in \mathcal{I}(p), \textit{ for } p \in \mathbf{P} \\
\mathcal{M}, a \models \neg\varphi & \textit{iff} & a \notin ||\varphi||_{\mathcal{M}} \\
\mathcal{M}, a \models \varphi_1 \wedge \varphi_2 & \textit{iff} & a \in ||\varphi_1||_{\mathcal{M}} \cap ||\varphi_2||_{\mathcal{M}} \\
\mathcal{M}, a \models \langle\leftarrow\rangle\varphi & \textit{iff} & a \in \{b \mid \exists c : b \leftarrow c \ \&\ c \in ||\varphi||_{\mathcal{M}}\} \\
\mathcal{M}, a \models \mu p.\varphi(p) & \textit{iff} & a \in \bigcap\{X \in 2^A \mid ||\varphi||_{\mathcal{M}[p:=X]} \subseteq X\}
\end{array}$$

where $||\varphi||_{\mathcal{M}[p:=X]}$ denotes the truth-set of φ once $\mathcal{I}(p)$ is set to be X. As usual, we say that: φ is valid in an argumentation model $\mathcal{M}$ iff it is satisfied in all pointed models of $\mathcal{M}$, i.e., $\mathcal{M} \models \varphi$; φ is valid in a class $\mathfrak{M}$ of argumentation models iff it is valid in all its models, i.e., $\mathfrak{M} \models \varphi$. All definitions are naturally generalizable to sets of formulae Φ.

5.1.3 Axiomatics.

The standard axiomatics for the μ-calculus built on modal system **K** suffices for our purposes. Logic K^μ is axiomatized by the following rules and axiom schemata.

(Prop)	propositional schemata
(K)	$[\leftarrow](\varphi_1 \rightarrow \varphi_2) \rightarrow ([\leftarrow]\varphi_1 \rightarrow [\leftarrow]\varphi_2)$
(Fixpoint)	$\varphi(\mu p.\varphi(p)) \leftrightarrow \mu p.\varphi(p)$
(MP)	IF $\vdash \varphi_1 \rightarrow \varphi_2$ AND $\vdash \varphi_1$ THEN φ_2
(N)	IF $\vdash \varphi$ THEN $\vdash [\leftarrow]\varphi$
(Least)	IF $\vdash \varphi_1(\varphi_2) \rightarrow \varphi_2$ THEN $\vdash \mu p.\varphi_1(p) \rightarrow \varphi_2$

So, the axiomatics of K^{μ} consists of the axiom system **K** axiomatizing $\langle\leftarrow\rangle$ plus schema `Fixpoint` and rule `Least`. Let us have a closer look at what they state. Axiom `Fixpoint` just states that $\mu p.\varphi(p)$ is indeed a fixpoint since a further application of φ still yields $\mu p.\varphi(p)$ and vice versa. Instead, rule `Least` guarantees that $\mu p.\varphi(p)$ is in fact the least fixpoint by imposing that if φ_2 is provably a pre-fixpoint of φ_1, then $\mu p.\varphi_1(p)$ provably implies φ_2.

5.1.4 Meta-theoretical results.

We list some relevant known results.

- Logic K^{μ} is sound and complete for the class $\mathfrak{A}$ of all argumentation models under the semantics given in Definition 4 [23]. Notice however that, unlike K^{-1} and K^{U}, the given axiomatics of K^{μ} is not strongly complete since it is obviously not compact.

- The satisfiability problem of K^{μ} is decidable [20].

- The complexity of the model-checking problem for K^{μ} is known to be in NP ∩ co-NP [13], however, it is still an open question whether it is in P.

The next result deserves some highlighting. First define the *alternation depth* of a formula of $\mathcal{L}^{\mathsf{K}^{\mu}}$ as the maximum number of μ/ν in a chain of nested fixpoints.

Fact 5 (Model-checking K^{μ})**.** *The complexity of the model-checking problem for a formula of size m and alternation depth d on a system of size n is* $O(m \cdot n^{d+1})$.

Proof. The result is proven in [10]. □

Such result will be used to establish the complexity of model-checking grounded extensions.

5.2 Grounded extensions in K^{μ}

The notion of grounded extension can be given a modal formulation within $\mathcal{L}^{\mathsf{K}^{\mu}}$. According to [9], the grounded extension of an argumentation framework $\mathcal{A}$ is the smallest complete extension. In an argumentation model $\mathcal{M}$, it is therefore a formula φ such that $\mathcal{M} \models (\varphi \rightarrow [\leftarrow]\neg\varphi) \wedge (\varphi \leftrightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$ and its truth-set $||\varphi||$ is the smallest among all other such formulae. In other words, φ is a formula whose truth-set is smallest among all the formulae which are conflict-free and which are a $[\leftarrow]\langle\leftarrow\rangle$-fixpoint. However, being the smallest $[\leftarrow]\langle\leftarrow\rangle$-fixpoint—which exists by Corollary 1—implies being conflict-free.

Theorem 4 (The least $[\leftarrow]\langle\leftarrow\rangle$-fixpoint is conflict-free)**.** *The following formula is a validity of* K^{μ}*:*

$$\mu p.[\leftarrow]\langle\leftarrow\rangle p \rightarrow [\leftarrow]\neg(\mu p.[\leftarrow]\langle\leftarrow\rangle p) \tag{24}$$

Proof. We proceed per absurdum. Take an argumentation model $\mathcal{M}$ such that $\mathcal{M} \models \mu p.[\leftarrow]\langle\leftarrow\rangle p \wedge \neg[\leftarrow]\neg(\mu p.[\leftarrow]\langle\leftarrow\rangle p)$. By the Definition 4 we obtain that $\mathcal{M} \models \mu p.[\leftarrow]\langle\leftarrow\rangle p$ and that there exist a arguments a, b such that $a \leftarrow b$ and $\mathcal{M}, b \models \mu p.[\leftarrow]\langle\leftarrow\rangle p$ while also $\mathcal{M}, a \models \mu p.[\leftarrow]\langle\leftarrow\rangle p$. We distinguish two cases: 1) there exists a finite chain $(a \leftarrow b \leftarrow b_1 \leftarrow \ldots \leftarrow b_n)$ of successors starting from a; 2) there exists an infinite such chain. If 1) is the case, then $\mathcal{M}, b_n \models$

$[\leftarrow]\varphi$ for any φ. Since both $\mathcal{M}, a \models \mu p.[\leftarrow]\langle\leftarrow\rangle p$ and $\mathcal{M}, b \models \mu p.[\leftarrow]\langle\leftarrow\rangle p$, then $\mathcal{M}, b_{n-1} \models \mu p.[\leftarrow]\langle\leftarrow\rangle p$ which, by Definition 4, means that for any p such that $\|[\leftarrow]\langle\leftarrow\rangle p\|_{\mathcal{M}} \subseteq \|p\|_{\mathcal{M}}$, $\mathcal{M}, b_{n-1} \models [\leftarrow]\langle\leftarrow\rangle p$, which is impossible given that for any φ $\mathcal{M}, b_n \models [\leftarrow]\varphi$ and hence that $\mathcal{M}, b_{n-1} \models \langle\leftarrow\rangle[\leftarrow]\neg p$. If 2) is the case, then we show that $\|\mu p.[\leftarrow]\langle\leftarrow\rangle p\|_{\mathcal{M}} = \emptyset$. This is the case since the two following sets are both pre-fixpoints but they have empty intersection: $\{c \in A \mid a \leftarrow^{2m} c\}$ and $\{c \in A \mid b \leftarrow^{2m} c\}$ where $\leftarrow^{2m}$ denotes reachability via $\leftarrow$ in an even number of steps. We thus obtain a contradiction. □

Just like Formulae 20 and 21 are theorems/validities of K^{U} (Theorems 2 and 3), so is Formula 24 a theorem/validity of K^{μ}. Again, we thus obtain a formalization of basic results of argumentation theory in a modal logic.

From Theorem 4 it follows that Formula 23 is a modal logic formulation of the notion of grounded extension. A grounded extension is the least $[\leftarrow]\langle\leftarrow\rangle$-fixpoint and, by Fact 4, it denotes a conflict-free set of arguments. We have the following result.

Theorem 5 (Model-checking grounded extensions)**.** *Given an argumentation model $\mathcal{M}$, it can be decided in polynomial time whether an argument a belongs to the grounded extension of $\mathcal{M}$, that is, whether $\mathcal{M}, a \models \mu p.[\leftarrow]\langle\leftarrow\rangle p$.*

Proof. Since $\mu p.[\leftarrow]\langle\leftarrow\rangle p$ has alternation depth 0, by Fact 5, it follows that model-checking $\mu p.[\leftarrow]\langle\leftarrow\rangle p$ can be done in $O(m \cdot n)$ where m is the size of $\mu p.[\leftarrow]\langle\leftarrow\rangle p$ and n the size of $\mathcal{M}$. □

5.3 Preferred extensions in K^{μ}?

At the other extreme of grounded extensions, are preferred extensions. In [9], preferred extensions are defined as maximal, with respect to set-inclusion, complete extensions. In this case, $\mathcal{L}^{\mathsf{K}^{\mu}}$ is not able to express such a notion since the greatest $[\leftarrow]\langle\leftarrow\rangle$-fixpoint is not necessarily conflict-free. In other words, and not unsurprisingly, we do not have the equivalent of Fact 4 for ν. To appreciate this, consider the 2 arguments cycle $\mathcal{A} = (\{a, b\}, \{(a, b), (b, a)\})$. In this case $\|\nu p.[\leftarrow]\langle\leftarrow\rangle p\|_{\mathcal{M}} = \{a, b\}$ which is obviously not conflict-free for any model $\mathcal{M}$. The point is that preferred extensions maximize acceptability, i.e., $[\leftarrow]\langle\leftarrow\rangle$, together with conflict-freeness, i.e. $\neg\langle\leftarrow\rangle$.

So, given an argumentation model $\mathcal{M}$, a formula φ is a preferred extension for $\mathcal{M}$ if and only if:

1. $\mathcal{M} \models Adm(\varphi)$, and
2. $\forall X \subseteq A$ if $\mathcal{M} \models Adm(X)$ and $\mathcal{M} \models \varphi \rightarrow X$ then $\mathcal{M} \models X \rightarrow \varphi$.

Clearly, item 2 involves a monadic second-order quantification. Abusing notation, if we had to put it into an extension of K^{U} with Π^1_1 quantification, we would obtain a formula like this:

$$\forall X((Adm(X) \wedge [\mathsf{U}](\varphi \rightarrow X)) \rightarrow [\mathsf{U}](X \rightarrow \varphi)) \tag{25}$$

with X not occurring in φ. That is, for any set X, if X is an admissible set and it contains $\|\varphi\|$, then $\|\varphi\|$ contains X. In other words, $\|\varphi\|$ is maximal among the admissible sets. Notice that Formula 25 denotes, like in the case of complete and stable extensions, a global property. Had Formula 25 to be properly formulated in Monadic Second Order Logic (MSO), we would obtain a much less succint formula.

6 Dialogue games via semantic games

The proof-theory of abstract argumentation is commonly given in terms of dialogue games [18]. The present section shows how modal semantics supports a general setting for the development of proof procedures based on games [15]. In particular we will focus on the so-called *evaluation games* or *model-checking games* where a proponent or verifier ($\exists$ve) tries to prove that a given formula φ holds in a point a of a model $\mathcal{M}$, while an opponent or falsifier ($\forall$dam) tries to disprove it.

The present section will describe the evaluation game for K^{U} which is a straightforward extension of the evaluation game for K but which, to the best of our knowledge, has not yet been investigated. For an exposition of evaluation games for K^{μ} we refer the reader to [22].

6.1 Evaluation game for K^{U}

We now introduce the game-theoretical semantics [15] of logic K^{U} placing it in the context of abstract argumentation. The notation is borrowed from [22].

Such a game is a *graph game*, that is, a game played by two agents on a directed graph, where each node—called position—is labelled by the player that is supposed to move next. The structure of the graph determines which are the *admissible moves* at any given position. If a player has to move in a certain position but there are no available moves, then it loses and its opponent wins. In general, graph games might have infinite paths, but this is not the case in the game we are going to introduce. A match of a graph game is then just the set of positions visited during play, that is, a complete path through the graph. Here is the formal definition of the evaluation game for K^{U}.

Definition 5 (Evaluation game for K^{U})**.** *Given a formula $\varphi \in \mathcal{L}^{\mathsf{K}^{\mathsf{U}}}$, and an argumentation model $\mathcal{M}$, the* evaluation game *$\mathcal{E}(\varphi, \mathcal{M})$ is defined by the following items.*

Players: *The set of players is $\{\exists, \forall\}$. An element from $\{\exists, \forall\}$ will be denoted P and its opponent $\overline{P}$*

Game form: *The* game form *of $\mathcal{E}(\varphi, \mathcal{M})$ is defined by following board game:*

Position	Turn	Available moves
$(\varphi_1 \vee \varphi_2, a)$	$\exists$	$\{(\varphi_1, a), (\varphi_2, a)\}$
$(\varphi_1 \wedge \varphi_2, a)$	$\forall$	$\{(\varphi_1, a), (\varphi_2, a)\}$
$(\langle\leftarrow\rangle\varphi, a)$	$\exists$	$\{(\varphi, b) \mid (a, b) \in\rightarrow^{-1}\}$
$([\leftarrow]\varphi, a)$	$\forall$	$\{(\varphi, b) \mid (a, b) \in\rightarrow^{-1}\}$
$(\langle\mathsf{U}\rangle\varphi, a)$	$\exists$	$\{(\varphi, b) \mid b \in A\}$
$([\mathsf{U}]\varphi, a)$	$\forall$	$\{(\varphi, b) \mid b \in A\}$
$(\bot, a)$	$\exists$	$\emptyset$
$(\top, a)$	$\forall$	$\emptyset$
(p, a) & $a \notin \mathcal{I}(p)$	$\exists$	$\emptyset$
(p, a) & $a \in \mathcal{I}(p)$	$\forall$	$\emptyset$
$(\neg p, a)$ & $a \in \mathcal{I}(p)$	$\exists$	$\emptyset$
$(\neg p, a)$ & $a \notin \mathcal{I}(p)$	$\forall$	$\emptyset$

Winning conditions: *Player P wins if and only if $\overline{P}$ has to play in a position with no available moves.*

Instantiation: *The* instance *of $\mathcal{E}(\varphi, \mathcal{M})$ with starting point (φ, a) is denoted $\mathcal{E}(\varphi, \mathcal{M})@(\varphi, a)$.*

The important thing to notice is that positions of the game are pairs of a formula and an argument, and that the type of formula in the position determines which player has to play: $\exists$ if the formula is a disjunction, a box, a false atom or $\bot$, and $\forall$ in the remaining cases.[5]

We can now define the notions of winning strategies and positions.

Definition 6 (Winning strategies and positions)**.** *A strategy for player P in an instantiated game $\mathcal{E}(\varphi, \mathcal{M})@(\varphi, a)$ is a function telling P what to do in any match played from position (φ, a). Such a strategy is* winning for *P if and only if, in any match played according to the strategy, P wins. A position (φ, a) in $\mathcal{E}(\varphi, \mathcal{M})$ is* winning for *P if and only if P has a winning strategy in $\mathcal{E}(\varphi, \mathcal{M})@(\varphi, a)$. The set of winning positions of $\mathcal{E}(\varphi, \mathcal{M})$ is denoted $Win_P(\mathcal{E}(\varphi, \mathcal{M}))$.*

From the point of view of game theory [17], the game described in Definition 5 and with the winning conditions introduced in Definition 6 is a two-players zero-sum game. Such games have the property that P wins if and only if

[5] Notice also that the game considers only positions consisting of formulae in positive normal form, that is, formulae where all negations are pushed inwards and occur only in front of atomic formulae.

$\overline{P}$ looses (zero-sum), and that they are determined, that is, each match has a winner [24].

It now remains to be proven that the game just introduced is adequate with respect to the semantics of K^{U}. To put it otherwise, we have to prove that if $\exists$ always wins then the formula defining the game is true at the point of instantiation, and that if a formula is true at a point in a model, then $\exists$ always wins the corresponding game instantiated at that point.

Theorem 6 (Adequacy of the evaluation game for K^{U})**.** *Let $\varphi \in \mathcal{L}^{\mathsf{K}^{\mathsf{U}}}$, and let $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ be an argumentation model. Then, for any argument $a \in A$, it holds that:*

$$(\varphi, a) \in Win_{\exists}(\mathcal{E}(\varphi, \mathcal{M})) \Longleftrightarrow \mathcal{M}, a \models \varphi.$$

Proof. We proceed by induction on the length l of φ.
Base. $l = 0$. We have four cases:

- $\varphi = \top$. Straightforward since (φ, a) is then always a winning position for $\exists$.
- $\varphi = \bot$. Straightforward since (φ, a) is then never a winning position for $\exists$.
- $\varphi = p$. It follows that if $a \in \mathcal{I}(p)$ then (φ, a) is a winning position for $\exists$ and if $a \notin \mathcal{I}(p)$ then (φ, a) is not a winning position for $\exists$.
- $\varphi = \neg p$. The converse argument applies.

Step. $l > 0$. The induction hypothesis is that for any subformula ψ of φ of length $l - 1$, and for any $b \in A$, $(\psi, b) \in Win_{\exists}(\mathcal{E}(\psi, \mathcal{M})) \Longleftrightarrow \mathcal{M}, b \models \psi$. We have the following cases:

- $\varphi = \psi_1 \wedge \psi_2$. From left to right. Assume $(\varphi, a) \in Win_{\exists}(\mathcal{E}(\varphi, \mathcal{M}))$. Now, φ is a conjunction, hence it is $\forall$'s turn to move. It follows that (ψ_1, a) and (ψ_2, a) are both winning positions for $\exists$ in the corresponding games. By induction hypothesis, we thus have $\mathcal{M}, a \models \psi_1$ and $\mathcal{M}, a \models \psi_2$. From right to left. Assume $\mathcal{M}, a \models \varphi$. It follows that $\mathcal{M}, a \models \psi_1$ and $\mathcal{M}, a \models \psi_2$. By induction hypothesis we obtain that both (ψ_1, a) and (ψ_2, a) are winning positions for $\exists$, and thus so is (φ, a).
- $\varphi = \psi_1 \vee \psi_2$. From left to right. Assume $(\varphi, a) \in Win_{\exists}(\mathcal{E}(\varphi, \mathcal{M}))$. It is $\exists$'s turn to move, so one of (ψ_1, a) and (ψ_2, a) should be a winning position in the corresponding game. Assume WLOG it to be (ψ_1, a). By induction hypothesis it follows that $\mathcal{M}, a \models \psi_1$ and therefore $\mathcal{M}, a \models \varphi$. From right to left. Assume $\mathcal{M}, a \models \varphi$ and assume WLOG that $\mathcal{M}, a \models \psi_1$. By induction hypothesis we obtain that $(\psi_1, a) \in Win_{\exists}(\mathcal{E}(\psi_1, \mathcal{M}))$. Since φ is a disjunction, it is $\exists$'s turn to move and therefore we conclude $(\varphi, a) \in Win_{\exists}(\mathcal{E}(\varphi, \mathcal{M}))$.
- $\varphi = \langle \leftarrow \rangle \psi$. From left to right. Assume $(\varphi, a) \in Win_{\exists}(\mathcal{E}(\varphi, \mathcal{M}))$. It is $\exists$'s turn to move. It follows that there is a position (ψ, b) such that $a \leftarrow b$ and such that is a winning position for $\exists$. By induction hypothesis we conclude that $\mathcal{M}, b \models \psi$ and hence $\mathcal{M}, a \models \langle \leftarrow \rangle \psi$. From right to left. Assume $\mathcal{M}, a \models \varphi$. It follows that there exists b such that $a \leftarrow b$ and $\mathcal{M}, b \models \psi$. By induction hypothesis we have that $(\psi, b) \in Win_{\exists}(\mathcal{E}(\psi, \mathcal{M}))$. But it is $\exists$'s turn to move, hence we conclude $(\varphi, a) \in Win_{\exists}(\mathcal{E}(\varphi, \mathcal{M}))$.

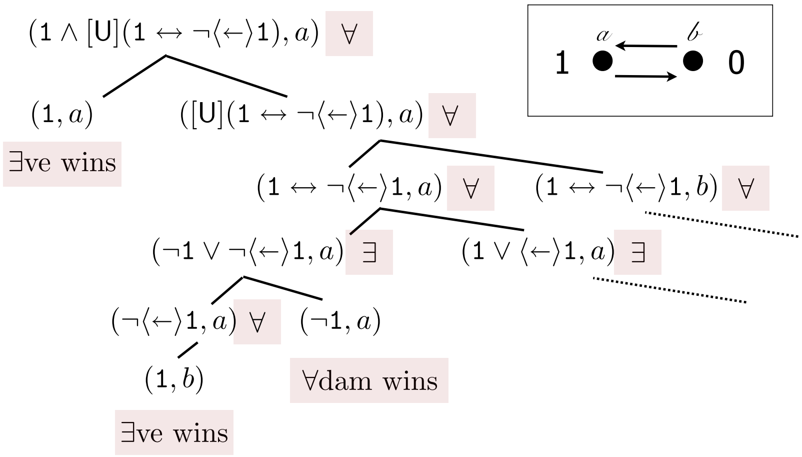
$(1 \wedge [\mathsf{U}](1 \leftrightarrow \neg\langle\leftarrow\rangle 1), a)$ ∀
$(1, a)$
∃ve wins
$([\mathsf{U}](1 \leftrightarrow \neg\langle\leftarrow\rangle 1), a)$ ∀
$(1 \leftrightarrow \neg\langle\leftarrow\rangle 1, a)$ ∀
$(1 \leftrightarrow \neg\langle\leftarrow\rangle 1, b)$ ∀
$(\neg 1 \vee \neg\langle\leftarrow\rangle 1, a)$ ∃
$(1 \vee \langle\leftarrow\rangle 1, a)$ ∃
$(\neg\langle\leftarrow\rangle 1, a)$ ∀
$(\neg 1, a)$
$(1, b)$
∀dam wins
∃ve wins
a
b
1
0

Figure 1: Evaluation game for stable extensions in the 2-cycle.

- $\varphi = [\leftarrow]\psi$. From left to right. Assume $(\varphi, a) \in Win_{\exists}(\mathcal{E}(\varphi, \mathcal{M}))$. It is ∀'s turn to move. It follows that for all $b \in A$ such that $a \leftarrow b$ $(\psi, b) \in Win_{\exists}(\mathcal{E}(\psi, \mathcal{M}))$. From this, by induction hypothesis, we conclude that for all $b \in A$ such that $a \leftarrow b$, $\mathcal{M}, b \models \psi$. From right to left. Assume $\mathcal{M}, a \models \varphi$. It follows that for all $b \in A$ such that $a \leftarrow b$, $\mathcal{M}, b \models \psi$. By induction hypothesis we thus obtain that for all $b \in A$, $(\psi, b) \in Win_{\exists}(\mathcal{E}(\psi, \mathcal{M}))$. This proves that $(\varphi, a) \in Win_{\exists}(\mathcal{E}(\varphi, \mathcal{M}))$.
- $\varphi = \langle \mathsf{U} \rangle \psi$. Similar to the case for $\varphi = \langle\leftarrow\rangle\psi$.
- $\varphi = [\mathsf{U}]\psi$. Similar to the case for $\varphi = [\leftarrow]\psi$.

This completes the proof. □

In the next section we illustrate how this type of semantic games can be used as a general setting for games checking whether an argument of a given framework belongs to a specific extension under a given labeling.

6.2 Games for model-checking extensions

The following example shows how the game-theoretical semantics of modal logic can be used to provide games for abstract argumentation. We choose to discuss in detail the game for stable semantics, which has remained an open question among argumentation theorists for a while [6]. Such a game neatly follows as a special case of the evaluation game for K^{U} .

Example 3 (Game for stable extensions)**.** *Consider the simple argumentation framework $\mathcal{A} = (\{a, b\}, \{(a, b), (b, a)\})$ consisting of two arguments a and b attacking each*

Adm :	$\mathcal{E}(\varphi \wedge [\mathsf{U}](\varphi \rightarrow ([\rightarrow]\neg\varphi \wedge [\leftarrow]\langle\leftarrow\rangle\varphi)), \mathcal{M})@(\varphi \wedge [\mathsf{U}](\varphi \rightarrow ([\rightarrow]\neg\varphi \wedge [\leftarrow]\langle\leftarrow\rangle\varphi), a)$
Complete :	$\mathcal{E}(\varphi \wedge [\mathsf{U}](\varphi \leftrightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)), \mathcal{M})@(\varphi \wedge [\mathsf{U}](\varphi \leftrightarrow [\leftarrow]\langle\leftarrow\rangle\varphi), a)$
Stable :	$\mathcal{E}(\varphi \wedge [\mathsf{U}](\varphi \leftrightarrow \neg\langle\leftarrow\rangle\varphi)), \mathcal{M})@(\varphi \wedge [\mathsf{U}](\varphi \leftrightarrow \neg\langle\leftarrow\rangle\varphi), a)$
Grounded :	$\mathcal{E}(\mu p.[\leftarrow]\langle\leftarrow\rangle p, \mathcal{M})@(\mu p.[\leftarrow]\langle\leftarrow\rangle p, a)$

Table 3: Games for admissible, complete, stable and grounded sets.

other, and consider the labeling $\mathcal{I}$ assigning $\mathtt{1}$ *to a and* $\mathtt{0}$ *to b (top right corner of Figure 6). We now want to run an evaluation game for checking whether a belongs to a stable extension corresponding to the truth-set of* $\mathtt{1}$*. Such game is the game $\mathcal{E}(\mathtt{1} \wedge Stable(\mathtt{1}), (\mathcal{A}, \mathcal{I}))$ initialized at position $(\mathtt{1} \wedge Stable(\mathtt{1}), a)$. That is, spelling out the definition of $Stable(\mathtt{1})$: $\mathcal{E}(\mathtt{1} \wedge [\mathsf{U}](\mathtt{1} \leftrightarrow \neg\langle\leftarrow\rangle\mathtt{1}))@(\mathtt{1} \wedge [\mathsf{U}](\mathtt{1} \leftrightarrow \neg\langle\leftarrow\rangle\mathtt{1}), a)$. Such a game, played according to the rules in Definitions 5 and 6, gives rise to the tree partially depicted in Figure 6.*

In the previous section and in the example we have focused only on logic K^{U}. However, logic K^{μ} can also be given an analogous game-theoretical semantics, which delivers the type of logic games necessary to check whether an argument a in a given model $\mathcal{M}$ belongs to the grounded extension $\mu p.[\leftarrow]\langle\leftarrow\rangle p$. We do not work out the details here and we refer the reader to [22].

In general, evaluation games permit us to give a systematic presentation of games for checking membership of an argument to admissible sets, as well as complete, stable and grounded extensions by instantiating a game $\mathcal{E}(\varphi, \mathcal{M})$ at the given argument where φ expresses the to-be-checked set or extension. Such systematization is provided in Table 3. Notice that what changes is precisely the formula defining the game.

Now the natural question arises of what is the precise relationship between the games just exposed and the dialogue games normally studied in the literature on argumentation theory (see, for instance, [18]). The next section is concerned with this question.

6.3 Model-checking games vs. dialogue games

Before closing the section it is worth looking at an essential difference between the type of games discussed here, and the dialogue, or discussion, games typically studied in argumentation theory. The best way to highlight such difference is by means of complexity-theoretic considerations.

We have shown, in the previous sections, that checking whether an argument belongs to *a specific* admissible set, or an extension (complete, stable or grounded) can be done in P. However, it is well-known in argumentation theory that checking whether an argument belongs to *an* extension (complete, stable or grounded) is an NP-complete problem. So where is the trick?

In model-checking (or evaluation) games you are given a model $\mathcal{M} = (\mathcal{A}, \mathcal{I})$, a formula φ and an argument a, and $\exists$ve is asked to prove that $\mathcal{M}, a \models \varphi$. In dialogue games, the check appointed to the proponent is inherently more complex since there, the input consists only of an argumentation framework $\mathcal{A}$, a formula φ and an argument a, and the proponent is asked to prove that there exists a labeling $\mathcal{I}$ such that $(\mathcal{A}, \mathcal{I}), a \models \varphi$. In modal logic terms, this is not a model-checking problem, but a satisfiability problem in a pointed frame which, in turn, is essentially a model-checking problem in MSO:

$$\mathcal{A} \models \forall p_1, \ldots, p_n \neg ST_a(\varphi) \tag{26}$$

where $p_1, \ldots, p_n$ are the atoms occurring in φ and $ST_a(\varphi)$ is the standard translation of φ realized in state a.[6]

To conclude, we might say that model-checking games provide a game-theoretical approach to the "easy" part of the more difficult problem tackled by dialogue games, that is, the tractable check that is done once a labeling is "guessed".

7 When are two arguments the same?

Since abstract argumentation neglects the internal structure of arguments, the natural question arises of when two arguments can be said to be the same, once such abstract perspective is assumed. Given that arguments are, after all, just "points" in a structure, the only way to compare them is to consider their "behavior" with respect to other arguments, that is, what they attack and by what are they attacked.

Studying such notion of "sameness" of arguments and argumentation frameworks is not just a mathematical diversion. A neat example where this issue appears in all its relevance is in legal reasoning, and in particular within common-law systems. Often, in such systems the so-called principle of *stare decisis* [16] holds. According to such a principle, a judge should rule cases that are "substantially the same" in the same way. However, since judicial cases can be profitably viewed as argumentation frameworks, being the same in this context seems to mean something like exhibiting the "same argumentative structure". In the present section we present a formal study of this simple intuition based on the logics introduced thus far, i.e., K^{U} and K^{μ}.

7.1 Sameness of arguments in K^{U}

The model-theoretic analysis of abstract argumentation exposed in the previous sections enable us with a well-investigated formalization of such a "behavioral equivalence" between points: bisimulation [1, 12]. It is well-known that logic K^{μ} is invariant under bisimulation. It is, in fact, the bisimulation-invariant fragment of MSOL [22]. In the present section we will focus on the specific notion of bisimulation which is tailored to K^{U}, also called *total bisimulation*.

We briefly recapitulate the notion of bisimulation [1, 12] presenting it in an argumentation-theoretic flavor.

Definition 7 (Bisimulation)**.** *Let $\mathcal{M} = (A, \rightarrow, \mathcal{I})$ and $\mathcal{M}' = (A', \rightarrow', \mathcal{I}')$ be two argumentation models. A bisimulation between $\mathcal{M}$ and $\mathcal{M}'$ is a non-empty relation $Z \subseteq A \times A'$ such that for any aZa':*

[6] For a definition of the standard translation we refer the reader to [1].

Atom: *a and a' are propositionally equivalent;*

Zig: *if $a \leftarrow b$ for some $b \in A$, then $a' \leftarrow b'$ for some $b' \in A'$ and bZb';*

Zag: *if $a' \leftarrow b'$ for some $b' \in A$ then $a \leftarrow b$ for some $\underline{i}nA$ and aZa'.*

A total bisimulation *is a bisimulation $Z \subseteq A \times A'$ such that its left projection covers A and its right projection covers A'. When a total bisimulation exists between $\mathcal{M}$ and $\mathcal{M}'$ we write $(\mathcal{M}, a) \leftrightarrows (\mathcal{M}', a')$.*

Now, since logic K^{U} is invariant under total bisimulation [1] and logic K^{μ} under bisimulation [12], we obtain a natural notion of "sameness" of arguments, which is weaker than the notion of isomorphism of argumentation frameworks. If two arguments are "the same" in this perspective, then they are equivalent from the point of view of argumentation theory, as far as the notions expressible in those logics are concerned. In particular, we obtain the following simple theorem for free.

Theorem 7 (Bisimilar arguments)**.** *Let $(\mathcal{M}, a)$ and $(\mathcal{M}', a')$ be two pointed argumentation models, and let Z be a total bisimulation between $\mathcal{M}$ and $\mathcal{M}'$. It holds that a belongs to the admissible set (complete extension, stable extension, grounded extension) φ if and only if a' belongs to the admissible set (complete extension, stable extension, grounded extension) φ.*

Proof. Follows directly from the fact that bisimulation implies K^{μ}-equivalence [12], and total bisimulation implies K^{U}-equivalence [1]. □

In other words, Theorem 7 states that if two arguments are totally bisimilar, then they are equivalent from the point of view of Dung's argumentation-theoretic semantics.

7.2 Total bisimulation games

We can associate a game to Definition 7. Such game checks whether two given pointed models $(\mathcal{M}, a)$ and $\mathcal{M}, a'$ are bisimular or not. The game is played by two players: **S**poiler, which tries to show that the two given pointed models are not bisimilar, and **D**uplicator which pursues the opposite goal. A match is started by **S**, then **D** responds, and so on. If and only if **D** moves to a position where the two pointed models are not propositionally equivalent, or if it cannot move, **S** wins. The following definition describes formally the game just sketched.

Definition 8 (Bisimulation game for K^{U})**.** *Given two pointed models $\mathcal{M}$ and $\mathcal{M}'$, the* total bisimulation game *$\mathcal{B}(\mathcal{M}, \mathcal{M}')$ is defined by the following items.*

Players: *The set of players is $\{\mathbf{D}, \mathbf{S}\}$. An element from $\{\mathbf{D}, \mathbf{S}\}$ will be denoted P and its opponent $\overline{P}$.*

Game form: *The* game form *of $\mathcal{B}(\mathcal{M}, \mathcal{M}')$ is defined by the following rule for available moves:*

Position	Available moves
$((\mathcal{M}, a)(\mathcal{M}', a'))$	$\{((\mathcal{M}, a)(\mathcal{M}', b')) \mid \exists b' \in A' : a' \leftarrow b'\}$
	$\cup\{((\mathcal{M}, b)(\mathcal{M}', a')) \mid \exists b \in A : a \leftarrow b\}$
	$\cup\{((\mathcal{M}, a)(\mathcal{M}', b')) \mid \exists b' \in A'\}$
	$\cup\{((\mathcal{M}, b)(\mathcal{M}', a')) \mid \exists b \in A\}$

Turn function: *If the round is even* **S** *plays, if it is odd* **D** *plays.*

Winning conditions: **S** *wins if and only if either D has moved to a position $((\mathcal{M}, a)(\mathcal{M}', a'))$ where a and a' do not satisfy the same labels, or* **D** *has no available moves. Otherwise* **D** *wins.*

Instantiation: *The* instance *of $\mathcal{B}(\mathcal{M}, \mathcal{M}')$ with starting position $((\mathcal{M}, a)(\mathcal{M}', a'))$ is denoted $\mathcal{B}(\mathcal{M}, \mathcal{M}')@(a, a')$.*

So, as we might expect, positions in a (total) bisimulation games are pairs of pointed models, that is, the pointed models that **D** tries to show are bisimilar. It might also be instructive to notice that such a game can have infinite matches, which, according to Definition 8 are thus won by **D**.

From Definition 8 we obtain the following notions of winning strategies and winning positions.

Definition 9 (Winning strategies and positions)**.** *A strategy for player P in an instantiated game $\mathcal{B}(\mathcal{M}, \mathcal{M}')@(a, a')$ is a function telling P what to do in any match played from position (a, a'). Such a strategy is* winning for *P if and only if, in any match played according to the strategy, P wins. A position $((\mathcal{M}, a)(\mathcal{M}', a'))$ in $\mathcal{B}(\mathcal{M}, \mathcal{M}')$ is* winning for *P if and only if P has a winning strategy in $\mathcal{B}(\mathcal{M}, \mathcal{M}')@(a, a')$. The set of all winning positions of game $\mathcal{B}(\mathcal{M}, \mathcal{M}')$ for P is denoted by $Win_P(\mathcal{B}(\mathcal{M}, \mathcal{M}'))$.*

Also in the case of (total) bisimulation games we obtain an adequacy theorem.

Theorem 8 (Adequacy of total bisimulation games)**.** *Take $(\mathcal{M}, a)$ and $(\mathcal{M}', a')$ to be two argumentation models. It holds that:*

$$((\mathcal{M}, a)(\mathcal{M}', a')) \in Win_{\mathbf{D}}(\mathcal{B}(\mathcal{M}, \mathcal{M}')) \Longleftrightarrow (\mathcal{M}, a) \rightleftarrows (\mathcal{M}', a').$$

Proof. The proof is standard and we refer the reader to [12]. □

In words, **D** has a winning strategy in the (total) bisimulation game $\mathcal{B}(\mathcal{M}, \mathcal{M}')@(a, a')$ if and only if $\mathcal{M}, a$ and $\mathcal{M}', a'$ are totally bisimilar. The following example illustrates how a total bisimulation game concretely looks like.

Example 4 (A total bisimulation game)**.** Consider two simple legal cases concerning the innocence or guiltiness of two defendants in two different trials. In the first one, two arguments a and b claiming the defendant to be guilty defeat

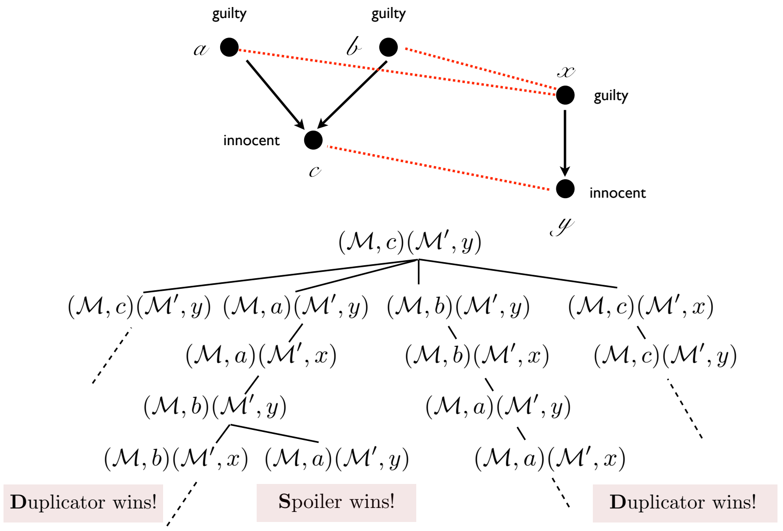
guilty
guilty
a
b
x
guilty
innocent
c
innocent
y
$(\mathcal{M},c)(\mathcal{M}',y)$
$(\mathcal{M},c)(\mathcal{M}',y)$ $(\mathcal{M},a)(\mathcal{M}',y)$ $(\mathcal{M},b)(\mathcal{M}',y)$ $(\mathcal{M},c)(\mathcal{M}',x)$
$(\mathcal{M},a)(\mathcal{M}',x)$ $(\mathcal{M},b)(\mathcal{M}',x)$ $(\mathcal{M},c)(\mathcal{M}',y)$
$(\mathcal{M},b)(\mathcal{M}',y)$ $(\mathcal{M},a)(\mathcal{M}',y)$
$(\mathcal{M},b)(\mathcal{M}',x)$ $(\mathcal{M},a)(\mathcal{M}',y)$ $(\mathcal{M},a)(\mathcal{M}',x)$
Duplicator wins!
Spoiler wins!
Duplicator wins!

Figure 2: Example of total bisimulation game.

an argument a claiming his/her innocence. In the second one, only one argument x claiming the defendant's guiltiness defeats an argument y for his/her innocence. The two argumentation models, $\mathcal{M}$ and $\mathcal{M}'$, are depicted at the top of Figure 2. A total bisimulation connects c with y, and a and b with x. Part of the extensive bisimulation game $\mathcal{B}(\mathcal{M},\mathcal{M}')@(c,y)$ is depicted in Figure 2. Notice that **D** wins on those infinite path where it can always duplicate **S**'s moves. On the other hand, it looses for instance when it replies to one of **S**'s moves $((\mathcal{M},b)(\mathcal{M}',y))$ by moving in the first model to state a which is labelled `guilty` while y is labelled `innocent`.

As the example suggests, to link bisimulation games to our intuitions, they can be viewed as an idealized version of the type of dialogues occurring in legal trials when old relevant cases are compared with new case at hands. The lawyer claiming their difference proceeds like the **S**poiler, while the lawyer claiming their equivalence, proceeds just like the **D**uplicator.

8 Conclusions and future work

The following is a non-exhaustive list of the future research lines we envision at the interface of modal logic and argumentation theory:

- Develop a study of preferred extensions in MSO along the line followed for K^{U} and K^{μ}.

- Check whether MSO is expressive enough to study semi-stable semantics [5].

- Investigate MSO model-checking games as a more appropriate logical setting for dialogue games than the modal model-checking games presented in the paper.
- Develop the application of the notion of bisimulation to the study of invariance in the context of argumentation theory, for instance by characterizing the notion of *accrual* within graded modal logic [8].
- Apply sabotage modal logic [21] to study the robustness of the membership of an argument to a certain set or extension denoted by a formula φ?

8.0.1 Acknowledgments.

This study has been funded under the NWO VENI grant 639.021.816. The author wishes to thank Sanjay Modgil for the inspiring conversation that sparkled this study.

References

[1] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001.

[2] P. Blackburn and J. van Benthem. Modal logic: A semantic perspective. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*, pages 1–84. Elsevier, 2006.

[3] P. Bradfield and C. Stirling. Modal mu-calculi. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*, pages 722–754. Elsevier, 2006.

[4] M. Caminada. On the issue of reinstatement in argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence. Proceedings of JELIA 2006*, pages 111–123, 2006.

[5] M. Caminada. Semi-stable semantics. In P. E. Dunne and T. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 121–130, 2006.

[6] M. Caminada and Y. Wu. An argument game for stable semantics. *Journal of the IGPL*, 17(1), 2009.

[7] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 1990.

[8] M. de Rijke. A note on graded modal logic. *Studia Logica*, 64(2):271–283, 2000.

[9] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[10] E. A. Emerson and C. Lei. Efficient model checking in fragments of the propositional mu-calculus. In *Proceedings of the 1st IEEE LICS*, pages 267–278, 1986.

[11] De Giacomo G. Eliminating 'converse' from converse PDL. *Journal of Logic, Language and Information*, 5(2):193–208, 1996.

[12] V. Goranko and M. Otto. Model theory of modal logic. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*, pages 249–329. Elsevier, 2007.

[13] E. Graedel and M. Otto. On logics with two variables. *Theoretical Computer Science*, 224:73–113, 1999.

[14] E. Hemaspaandra. The price of universality. *Notre Dame Journal of Formal Logic*, 37(2):174–203, 1996.

[15] J. Hintikka and G. Sandu. Game-theoretical semantics. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, chapter 6, pages 361–410. Elsevier, 1997.

[16] L. Kornhauser. An economic perspective on stare decisis. *Chicago-Kent Law Review*, 65:63–92, 1989.

[17] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

[18] H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. *Handbook of Philosophical Logic*, IV:218–319, 2002. Second Edition.

[19] K. Schild. A correspondence theory for terminological logics: preliminary report. In *Proceedings of IJCAI-91, 12th International Joint Conference on Artificial Intelligence*, pages 466–471, Sidney, AU, 1991.

[20] R. S. Streett and E. A. Emerson. An automata theoretic decision procedure for the propositional mu-calculus. *Information and Computation*, 81:249–264, 1989.

[21] J. van Benthem. An essay on sabotage and obstruction. In *Mechanizing Mathematical Reasoning*, LNCS, pages 268–276. Springer, 2005.

[22] Y. Venema. The modal mu-calculus. In *The 18th European Summer School in Logic, Language and Information*, Malaga, Spain, 31 July - 11 August, 2006 2006.

[23] I. Walukiewicz. Completeness of Kozen's axiomatization of the propositional mu-calculus. *Information and Computation*, 157:142–182, 2000.

[24] E. Zermelo. über eine anwendung der mengenlehre auf die theorie des schachspiels. In *Proceedings of the 5th Congress Mathematicians*, pages 501–504. Cambridge University Press, 1913.

$c_{\mathcal{A}}$ characteristic function of $\mathcal{A}$	iff	$c_{\mathcal{A}} : 2^A \longrightarrow 2^A$ s.t.
		$c_{\mathcal{A}}(X) = \{a \mid \forall b : [b \rightarrow a \Rightarrow \exists c \in X : c \rightarrow b]\}$
X conflict-free in $\mathcal{A}$	iff	$\nexists a, b \in X$ s.t. $a \rightarrow b$
X admissible set of $\mathcal{A}$	iff	$X \subseteq c_{\mathcal{A}}(X)$
	iff	X is a pre-fixpoint of $c_{\mathcal{A}}$
X complete extension of $\mathcal{A}$	iff	X is conflict-free and $X = c_{\mathcal{A}}(X)$
	iff	X is a conflict-free fixpoint of $c_{\mathcal{A}}$
X grounded extension of $\mathcal{A}$	iff	X is the minimal complete
	iff	X is the least fixpoint of $c_{\mathcal{A}}$
X preferred extension of $\mathcal{A}$	iff	X is a maximal complete
X stable extension of $\mathcal{A}$	iff	X is a complete and $\forall b \notin X, \exists a \in X : a \rightarrow b$
	iff	$X = \{a \in A \mid \nexists b \in X : b \rightarrow a\}$

Table 4: Basics of argumentation theory.

A Completeness of logic K^{-1}

Theorem 9 (Soundness and strong completeness of K^{-1})**.** *Logic K^{-1} is sound and strongly complete for the class $\mathfrak{A}$ of all argumentation models under the semantics given in Definition 2.*

Sketch of proof. Logic K^{-1} extends logic **K** with the **Conv** axiom. Logic **K** is defined on the sublanguage of $\mathcal{L}^{\mathsf{K}^{-1}}$ containing only one modality (either $\langle\rightarrow\rangle$ or $\langle\leftarrow\rangle$), and is sound and strongly complete with respect to $\mathfrak{A}$ [1]. To obtain the desired results it suffices to show that the canonical model of K^{-1} is such that $\langle\rightarrow\rangle$ is interpreted on the converse of the relation on which $\langle\leftarrow\rangle$ is interpreted, and vice versa. Let $\mathcal{M}^{\mathsf{K}^{-1}} = (A^{\mathsf{K}^{-1}}, R^{\mathsf{K}^{-1}}, \mathcal{I}^{\mathsf{K}^{-1}})$ be the canonical model of K^{-1}. We want to prove that, for all $a, a' \in A^{\mathsf{K}^{-1}}$: $aR^{\mathsf{K}^{-1}}a'$ if and only if $a'R^{\mathsf{K}^{-1}-1}a$. [Left to right] Assume $aR^{\mathsf{K}^{-1}}a'$ and suppose $\varphi \in a$. For axiom **Conv**, it follows that $[\rightarrow]\langle\leftarrow\rangle\varphi \in a$ and therefore, since $aR^{\mathsf{K}^{-1}}a'$, $\langle\leftarrow\rangle\varphi \in a'$. Hence, by the definition of the canonical accessibility relation, $a'R^{\mathsf{K}^{-1}-1}a$. [Right to left] An analogous argument applies. □

B Basics of argumentation theory

Let $\mathcal{A} = (A, \rightarrow)$ be an argumentation framework where A is a set of arguments and $\rightarrow \subseteq A \times A$. Table 8.0.1 briefly recapitulates the key notions developed in

[9] which are considered in the paper. For an explanation of the order-theoretic notions involved in the definitions we refer the reader to [7].

The notions in Table 8.0.1 obtain the following intuitive reading. The characteristic function assigns to each set of arguments X the set of arguments $c_{\mathcal{A}}(X)$ which X defends—by attacking all the attackers of $c_{\mathcal{A}}(X)$. The notion of conflict-freeness is self-explanatory. An admissible set is a set of arguments X which is able to defend all its attackers. By considering those admissible sets which contain all their defenders, we obtain the notion of complete extension, which somehow formalizes the idea of a 'reasonable' position in an argumentation, that is, a position which has no conflicts, and which consists exactly of all that it can successfully defend.

Stable, grounded and preferred extensions can all be considered to be refinements of this latter notion. A grounded extension, instead, represents what all complete extensions have in common. In a way, it formalizes the notion of what should be at least taken as 'reasonable' within the current argumentation. On the contrary, preferred extensions are maximal complete extensions which remain conflict-free and, as such, they represent somehow the most it can be 'reasonably' claimed within the given argumentation framework. Finally, a stable extension is a set of arguments X which is a complete extension and which attacks all arguments which do not belong to X itself. As such, it can be viewed as an 'aggressive' position within an argumentation.