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 games, bisimulation games); third, to import results (completeness of axiomatizations, complexity of model-checking, adequacy of games) from modal logic to argumentation theory.

Contents

*modalarg.tex, *Tuesday 14th December, 2010, 12:46.*

1 Introduction

The paper presents a study in the formal foundations of abstract argumentation theory as introduced in [11] by applying methods and techniques borrowed from modal logic [1]. The paper 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, adequacy) from modal logic to argumentation theory, and to do that essentially for free. Also, as it is often the case in the cross-fertilization of different formalisms, such perspective opens up interesting lines of research which were thus far hidden to the attention of argumentation theorists.

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 [11]. 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 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 presupposes some knowledge of modal logic as well as of argumentation theory. However, the latter is briefly recapitulated in Appendix A. The remainder of 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 [11], 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 [11], 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. Related work as well as gaps in the present state of this study are discussed in Section 8. Conclusions follow in Section 9 where future research lines are also sketched.

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. If an argumentation framework can be viewed as a Kripke frame, as explained in the introduction , then an argumentation framework plus a function assigning names from a set $\mathbf{P}$ to sets of arguments can be viewed as a Kripke model [1].

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. The following example applies this insight to argumentation labeling functions [5].

Example 1. *(Argument labelings as argumentation models) In argumentation theory, a labeling function [5] is a function* $l : \{\mathtt{1}, \mathtt{0}, ?\} \longrightarrow A$ *from the set of three labels* $\{ \mathtt{1}, \mathtt{0}, ? \}$*—intuitively in, out, undecided—to the set of arguments* A*.*

From a logical point of view, such a function is equivalent to a valuation function $\mathcal{I} : \mathbf{P} \longrightarrow 2^A$ *with the further constraint that each argument can get at most one label which, in propositional logic, amounts to the following formula:*

$$\mathtt{Label} := (\mathtt{1} \wedge \neg \mathtt{0} \wedge \neg ?) \vee (\neg \mathtt{1} \wedge \mathtt{0} \wedge \neg ?) \vee (\neg \mathtt{1} \wedge \neg \mathtt{0} \wedge ?).$$

As a consequence, a framework $\mathcal{A}$ *with a labeling function is nothing but an argumentation model* $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ *s.t.* $\mathcal{M} \models \mathtt{Label}$*. We will come back later to the sort of labeling used in argumentation theory to characterize extensions, and show that they can be characterized by modal formulae.*

Formula `Label` in the example is just a propositional formula but what is typically interesting in argumentation theory are statements of the sort: "argument a is attacked by an argument in a set φ"; "argument a is defended by the set φ", or, "φ attacks an attacker of argument a". These are modal statements, and in order to express them, it suffices to introduce a dedicated modal operator $\langle \leftarrow \rangle$ whose intuitive reading is "there exists an attacking argument such that". The next section introduces the kind of formal language needed for expressing them.

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):

$$(\mathcal{A}, \mathcal{I}), a \models \langle\rightarrow\rangle\varphi \quad (2)$$
$$(\mathcal{A}, \mathcal{I}), a \models \langle\leftarrow\rangle\varphi \quad (3)$$
$$(\mathcal{A}, \mathcal{I}), a \models \langle\rightarrow\rangle\langle\rightarrow\rangle\varphi \quad (4)$$

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*

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

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.

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 B for a the proof).
- The satisfiability problem of K^{-1} is P-reducible to the one of K in the presence of a background theory [14], which is known to be EXP-complete [22].

In the next section the logic just introduced is used to start off with a first 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 [11] 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 $\rightarrow$ and $\leftarrow$ for conflict-freeness)**.** *Let $\mathcal{M}$ be an argumentation model. It holds that:*

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

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:

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

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

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

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.

3.2 Complete and stable extensions

In [11], 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:*

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

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., $\{a \in A \mid \forall b \in A : [b \rightarrow a \Rightarrow \exists c \in X : c \rightarrow b]\}$.[2] Does logic K^{-1} have a syntactic counterpart of the characteristic function? The answer turns out to be yes.

Let $\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}$ be the language 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 $\mathfrak{ter}_{\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}} = (\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}, \wedge, \neg, \bot, [\leftarrow]\langle\leftarrow\rangle)$. Finally, let $\mathcal{I}^* : \mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle} \longrightarrow 2^A$ be the inductive extension of a valuation function $\mathcal{I} : \mathbf{P} \longrightarrow 2^A$

[2] It might be worth mentioning the following. Let $c_{\mathcal{A}}(A)$ be the set of images obtained by applying $c_{\mathcal{A}}$ to 2^A. It is easy to show that $\bigcup_{i=1}^{n} c_{\mathcal{A}}(X_i) = c_{\mathcal{A}}(\bigcup_{i=1}^{n} X_i)$ and its dual hold for $X_i \subseteq A$. So, $c_{\mathcal{A}}(A)$ forms a complete lattice of sets [9]. Such a lattice is also bounded by $c_{\mathcal{A}}(\emptyset)$ and A.

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

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

according to the semantics given in Definition 3. We can prove the following result.

Theorem 1 ($c_{\mathcal{A}}$ vs. $[\leftarrow]\langle\leftarrow\rangle$)**.** *Let $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ be an argumentation model. Function $\mathcal{I}^*$ is a homomorphism from $\mathtt{ter}_{\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}}$ 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. □

Theorem 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 [11]. 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, for any $\mathcal{M}$, $\mathcal{M}^s$:*

$$\begin{aligned}
\textbf{Monotonicity}: &\quad \mathcal{M} \models \varphi_1 \rightarrow \varphi_2 \Longrightarrow \mathcal{M} \models [\leftarrow]\langle\leftarrow\rangle\varphi_1 \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi_2 \\
\textbf{Normality}: &\quad \mathcal{M}^s \models \varphi \rightarrow \bot \Longrightarrow \mathcal{M}^s \models [\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow \bot
\end{aligned}$$

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.[3]

[3] 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 Argumentation in $\mathsf{K}^{\forall}$: 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 $\mathsf{K}^{\forall}$, that is, the minimal normal modal logic K extended with the universal modality.

4.1 Logic $\mathsf{K}^{\forall}$

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

4.1.1 Language.

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

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

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

Logic $\mathsf{K}^{\forall}$ 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"—$\langle\forall\rangle$—together with their duals.

4.1.2 Semantics.

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

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

$$\mathcal{M}, a \models \langle\leftarrow\rangle\varphi \quad \textit{iff} \quad \exists b \in A : (a, b) \in \rightarrow^{-1} \text{ AND } \mathcal{M}, b \models \varphi$$
$$\mathcal{M}, a \models \langle\forall\rangle\varphi \quad \textit{iff} \quad \exists b \in A : \mathcal{M}, b \models \varphi$$

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

4.1.3 Axiomatics.

The logic $\mathsf{K}^{\forall}$ is axiomatized as follows:

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

with $i \in \{\leftarrow, \forall\}$.

4.1.4 Meta-theoretical results.

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

- Logic $\mathsf{K}^{\forall}$ is sound and strongly complete for the class $\mathfrak{A}$ of argumentation frames [1, Ch. 7].
- The complexity of deciding whether a formula of $\mathcal{L}^{\mathsf{K}^{\forall}}$ is satisfiable is EXP-complete [17].
- The complexity of checking whether a formula of $\mathcal{L}^{\mathsf{K}^{\forall}}$ is satisfied by a pointed model $\mathcal{M}$ is P-complete [16].

4.2 Doing argumentation in $\mathsf{K}^{\forall}$

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 $\mathsf{K}^{\forall}$ is expressive enough to capture the following notions in the object-language.

$$Acc(\varphi, \psi) := [\forall](\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\psi) \tag{14}$$

$$CFree(\varphi) := [\forall](\varphi \rightarrow \neg\langle\leftarrow\rangle\varphi) \tag{15}$$

$$Adm(\varphi) := [\forall](\varphi \rightarrow ([\leftarrow]\neg\varphi \wedge [\leftarrow]\langle\leftarrow\rangle\varphi)) \tag{16}$$

$$Complete(\varphi) := [\forall]((\varphi \rightarrow [\leftarrow]\neg\varphi) \wedge (\varphi \leftrightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)) \tag{17}$$

$$Stable(\varphi) := [\forall](\varphi \leftrightarrow \neg\langle\leftarrow\rangle\varphi) \tag{18}$$

These definitions restate the meta-language definitions summarized in Tables 1 and 2. Let us explain them in details again. A set of arguments φ is acceptable with respect to the set of arguments ψ if and only all φ-arguments are such that for all their attackers there exists a defender in ψ (Formula 14). A set of arguments φ is conflict free if and only if all φ-arguments are such that none of their attackers is in φ (Formula 15). A set of arguments φ is admissible if and only if it is conflict free and acceptable with respect to itself (Formula 16). A set φ is a complete extension if and only if it is conflict free and it is equivalent to the set of arguments all the attackers of which are attacked by some φ-argument (Formula 17). Finally, a set φ is a stable extension if and only if it is equivalent to the set of arguments whose attackers are not in φ (Formula 18). The adequacy of these definitions with respect to the standard ones (see Table A in Appendix A) is easily checked.

Example 2. *(Argumentation labelings in $\mathsf{K}^{\forall}$) According to [5], a labeling function is a* complete labeling *if and only if the following holds for each argument: a) an argument is labeled* 1*, i.e., in, iff all its attackers are labeled* 0*, i.e., out. b) an argument is labeled* 0*, i.e., out, iff there exists at least one attacker labeled* 1*. The reformulation of a)-b) in $\mathsf{K}^{\forall}$ goes as follows:*

$$[\forall]((1 \leftrightarrow [\leftarrow]0) \wedge (0 \leftrightarrow \langle\leftarrow\rangle 1) \wedge \mathtt{Label}) \tag{19}$$

where Label *is the propositional formula described in Example 1. So, a valuation $\mathcal{I}$ on an alphabet containing* 1, 0 *and* ? *is a complete labeling for an argumentation framework $\mathcal{A}$ iff the model $(\mathcal{A}, \mathcal{I})$ satisfies Formula 19. Also, it is a matter of propositional reasoning to see that Formula 19 is equivalent to the following formula:*

$$Compl(1) \wedge [\forall]((0 \leftrightarrow \langle\leftarrow\rangle 1) \wedge \mathtt{Label}) \tag{20}$$

In words, this means that a function $\mathcal{I}$ on an alphabet containing 1*,* 0 *and* ? *is a complete labeling of $\mathcal{A}$ if and only if the model $(\mathcal{A}, \mathcal{I})$ makes* 1 *to be a complete extension (Formula 17) and evaluates the labels* 0 *and* ? *accordingly. We obtain therefore a direct correspondence between complete labelings and complete extensions. The same could be done for stable extensions.*

Logic $\mathsf{K}^{\forall}$ 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 [11], formalized and proved within $\mathsf{K}^{\forall}$.

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

$$Adm(\varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow Adm(\varphi \vee \psi) \wedge Acc(\xi, \varphi \vee \psi) \tag{21}$$

Sketch. A full formal derivation is given in Appendix C. □

Notice that Theorem 2 is, in fact, a generalized version of the Fundamental Lemma proven in [11]. It states that if φ is admissible and both ψ and ξ are acceptable with respect to it then also $\psi \vee \xi$ is admissible and ξ is acceptable with respect to $\varphi \vee \psi$.

We provide one more example of theorems of abstract argumentation which can be obtained as formal theorems of $\mathsf{K}^{\forall}$.

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

$$Stable(\varphi) \rightarrow Adm(\varphi) \tag{22}$$
$$Stable(\varphi) \rightarrow Complete(\varphi) \tag{23}$$

Proof. Formula 22 follows from Fact 2 and the completeness of $\mathsf{K}^{\forall}$. Formula 23 is a direct corollary of Formula 22, the definition of $Stable(\varphi)$, the definition of $Complete(\varphi)$ and the completeness of $\mathsf{K}^{\forall}$. □

Formulae 22 and 23 state well-known facts about the relative strength of admissible, complete and stable extensions. 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 $\mathsf{K}^{\forall}$, 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

The present section shows what kind of modal machinery is needed to capture the notion of grounded extension left aside in Section 2. In [11], the grounded extension is defined as the smallest fixpoint of the characteristic function of an argumentation framework (see Table A).

5.1 Characteristic functions and 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[4] 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 take the $\leftarrow$ fragment of K^{-1} and enhance it with fixpoint operators, thus moving into the realm of the so-called μ-calculi [4].

5.2 A μ-calculus for argumentation

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

5.2.1 Language.

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

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

[4]We refer the interested reader to [9] for a neat formulation of this result.

where p ranges over **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.[5] In general, the notation $\varphi(\psi)$ stands for ψ occurs in φ. The usual definitions for Boolean and modal operators can be applied. Intuitively, $\mu p.\varphi(p)$ denotes the smallest formula p such that $p \leftrightarrow \varphi(p)$. This intuition is made precise in the semantics of $\mathcal{L}^{\mathsf{K}^\mu}$ given in Definition 4. The greatest fixpoint operator ν can be defined from μ as follows: $\nu x.\varphi(x) := \neg\mu x.\neg\varphi(\neg x)$.

5.2.2 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:*

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

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.

We have now all the logical machinery in place to express the notion of grounded extension. Set $\varphi(p) := [\leftarrow]\langle\leftarrow\rangle p$, that is, take $\varphi(p)$ to be the modal version $[\leftarrow]\langle\leftarrow\rangle$ of the characteristic function, and apply it to formula p. What we obtain is a modal formula expressing the least fixpoint of a characteristic function, that is, the grounded extension:

$$Grounded := \mu p.[\leftarrow]\langle\leftarrow\rangle p \tag{24}$$

Notice that, unlike the notions formalized in Formulae 14-18, the grounded extension of a framework is always unique and does not depend on the particular labeling of a given model.

5.2.3 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

[5] 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.

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.

5.2.4 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 [27]. Notice however that, unlike K^{-1} and $\mathsf{K}^{\forall}$, the given axiomatics of K^μ is not strongly complete since it is obviously not compact.
- The satisfiability problem of K^μ is decidable [23].
- The complexity of the model-checking problem for K^μ is known to be in NP $\cap$ co-NP [16]. However, it is known that the model-checking problem for a formula of size m and alternation depth d on a system of size n is $O(m \cdot n^{d+1})$ [13], where the alternation depth of a formula of $\mathcal{L}^{\mathsf{K}^\mu}$ is the maximum number of $\mu/\neg\mu\neg$ alternations in a chain of nested fixpoints.

5.3 Doing argumentation in K^μ

Like in Section 4.2 we give a couple of examples of the kind of argumentation-theoretic results formalizable in K^μ.

Theorem 4 (Grounded extension is conflict-free)**.** *The following formula is a validity of* K^μ*:*

$$Grounded \rightarrow \neg[\leftarrow]Grounded \tag{25}$$

Proof. Consider Formula 24 and 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. □

Like Theorem 1, Theorem 4 provides a modal logic formulation of an argumentation-theoretic result.

As to the complexity of model-checking grounded extensions, it turns out to be tractable.

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 the result reported in Section 5.2.4, 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 $\mathcal{M}$. □

6 Dialogue games via semantic games

The proof-theory of abstract argumentation is commonly given in terms of dialogue games [21]. The present section shows how modal semantics supports a general setting for the development of proof procedures based on games [18]. In particular we will focus on the so-called *evaluation games* or *model-checking games* where a proponent or verifier (∃ve) tries to prove that a given formula φ holds in a point a of a model $\mathcal{M}$, while an opponent or falsifier (∀dam) tries to disprove it.

The present section will describe the evaluation game for $\mathsf{K}^{\forall}$ 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 [26].

6.1 Evaluation game for $\mathsf{K}^{\forall}$

We now introduce the game-theoretical semantics [18] of logic $\mathsf{K}^{\forall}$ placing it in the context of abstract argumentation. The notation is borrowed from [26].

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 $\mathsf{K}^{\forall}$.

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

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

Table 3: Rules of the evaluation game for $\mathsf{K}^{\forall}$.

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 the rules given in Table 3.*

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 game* $\mathcal{E}(\varphi, \mathcal{M})$ *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.[6]

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*

[6]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.

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 [20], 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 $\overline{P}$ looses (zero-sum), and that they are determined, that is, each match has a winner [28].

It now remains to be proven that the game just introduced is adequate with respect to the semantics of $\mathsf{K}^{\forall}$. 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 $\mathsf{K}^{\forall}$)**.** *Let $\varphi \in \mathcal{L}^{\mathsf{K}^{\forall}}$, 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 $(\varphi, a) \in Win_{\exists}(\mathcal{E}(\varphi, \mathcal{M}))$.

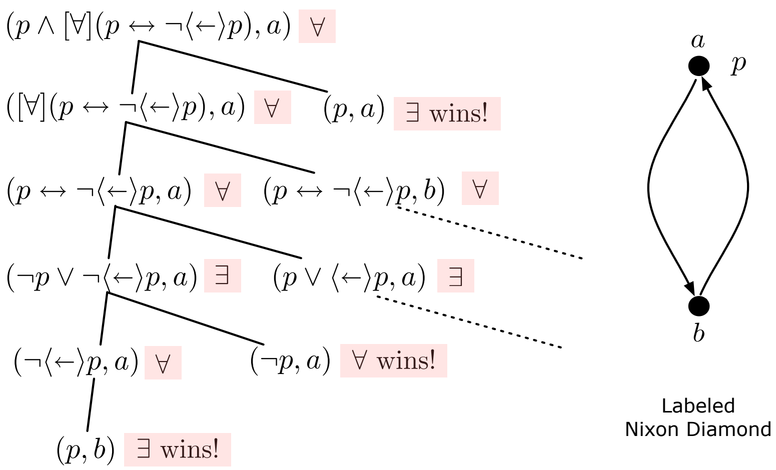

Figure 1: Game for stable extensions in the 2-cycle with labeling (valuation) function.

- $\varphi = \langle\leftarrow\rangle\psi$. From left to right. Assume $(\varphi, a) \in Win_\exists(\mathcal{E}(\varphi, \mathcal{M}))$. It is ∃'s turn to move. It follows that there is a position (ψ, b) such that $a \leftarrow b$ and such that is a winning position for ∃. 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 ∃'s turn to move, hence we conclude $(\varphi, a) \in Win_\exists(\mathcal{E}(\varphi, \mathcal{M}))$.

- $\varphi = [\leftarrow]\psi$. 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 $(\varphi, a) \in Win_\exists(\mathcal{E}(\varphi, \mathcal{M}))$.

- $\varphi = \langle\forall\rangle\psi$. Similar to the case for $\varphi = \langle\leftarrow\rangle\psi$.

- $\varphi = [\forall]\psi$. 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 details the game for stable semantics, which has remained an open

Adm :	$\mathcal{E}(\varphi \wedge Adm(\varphi), \mathcal{M})@(\varphi \wedge Adm(\varphi), a)$
Complete :	$\mathcal{E}(\varphi \wedge Complete(\varphi), \mathcal{M})@(\varphi \wedge Complete(\varphi), a)$
Stable :	$\mathcal{E}(\varphi \wedge Stable(\varphi)), \mathcal{M})@(\varphi \wedge Stable(\varphi), a)$
Grounded :	$\mathcal{E}(Grounded, \mathcal{M})@(Grounded, a)$

Table 4: Games for admissible, complete, stable and grounded sets.

question among argumentation theorists for a while [8]. Such a game neatly follows as the evaluation game for formula *Stable* (Formula 18) of $\mathsf{K}^{\forall}$.

Example 3 (Model-checking the Nixon diamond)**.** *Let $\mathcal{A} = (\{a, b\}, \{(a, b), (b, a)\})$ be an argumentation framework consisting of two arguments a and b attacking each other (i.e., the Nixon diamond), and consider the labeling $\mathcal{I}$ assigning* $\mathtt{1}$ *to a and* $\mathtt{0}$ *to b (top right corner of Figure 1). 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 $Stable(\mathtt{1})$: $\mathcal{E}(\mathtt{1} \wedge [\forall](\mathtt{1} \leftrightarrow \neg\langle\leftarrow\rangle\mathtt{1}))@(\mathtt{1} \wedge [\forall](\mathtt{1} \leftrightarrow \neg\langle\leftarrow\rangle\mathtt{1}), a)$. Such a game, played according to the rules in Definitions 5 and 6, gives rise to the tree partially depicted in Figure 1.*

In the previous section and in the example we have focused only on logic $\mathsf{K}^{\forall}$. 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 [26].

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 4. Notice that what changes is only the modal formula inputted in 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, [21]). The next section is concerned with this question.

6.3 Model-checking games vs. dialogue games

The best way to highlight the difference between model-checking games and dialogue games is by pointing considerations of a complexity-theoretic kind. We have seen, in Sections 4.2 and 5.3, that checking whether an argument belongs to a specific admissible set, or an extension (complete, stable or grounded) can be done in polynomial time. However, it is well-known that checking whether an argument belongs to an extension can be harder (e.g. NP-complete for stable extensions [12]). So where is the trick?

In model-checking games you are given a model $\mathcal{M} = (\mathcal{A}, \mathcal{I})$, a formula φ and an argument a, and ∃ve is asked to prove that $\mathcal{M}, a \models \varphi$. In dialogue games, the check appointed to ∃ve is inherently more complex since the input consists there of only an argumentation framework $\mathcal{A}$, a formula φ and an argument a. ∃ve is then asked to prove that there exists a labeling $\mathcal{I}$ such that $(\mathcal{A}, \mathcal{I}), a \models \varphi$. This is not a model-checking problem but a satisfiability problem in a pointed frame [1] which, in turn, is essentially a model-checking problem in monadic second-order logic: "$\mathcal{A} \models \forall p_1, \ldots, p_n \neg ST_a(\varphi)$?" 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.[7]

To conclude, we might say that the games defined above provide a proof procedure for a reasoning task which is computationally simpler than the one tackled by standard dialogue games. It should be noted, however, that this is no intrinsic limitation to the logic-based approach advocated in the present paper. Model-checking games for monadic second-order logic (or rather for appropriate fragments of it) would accommodate dialogue games in their entirety, lifting the sort of systematization they enable—in the form exemplified by Table 4—to dialogue games.

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 equivalent, or be "the same", from the point of view of argumentation theory. Such a notion of equivalence will necessarily be of a structural nature and, to be of any interest, be weaker than plain isomorphism. The study of a notion of equivalence for argumentation has not received attention yet by the argumentation theory community, except for one recent notable exception [19], which defines a notion of strong equivalence for argumentation frameworks, borrowed from the analogous notion developed in logic programming.

Modal logic offers a readily available notion of structural equivalence, the notion of bisimulation (with all its variants) [1, 15]. This section sketches the use of bisimulation for argumentation theoretic purposes. To illustrate the issue we use a simple motivating example depicted in Figure 2. We have two labelled argumentation frameworks which both contain an argument labeled p which is attacked by some arguments labelled q. Now the question would be: are the two p-arguments equivalent as far as abstract argumentation theory is concerned? The answer is yes, and the next sections explain why.

7.1 Indistinguishability of arguments in $\mathsf{K}^{\forall}$

It is well-known that logic K^{μ} is invariant under bisimulation. It is, in fact, the bisimulation-invariant fragment of monadic second-order logic [26]. In the present section we will focus on the specific notion of bisimulation which is tailored to $\mathsf{K}^{\forall}$, also called *total bisimulation*.

We briefly recapitulate the notion of bisimulation [1, 15] presenting it in an argumentation-theoretic flavor.

[7] For the standard translation we refer the reader to [1].

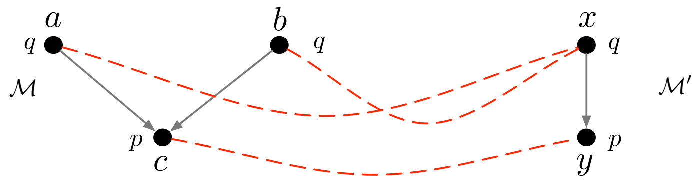

Figure 2: Two (totally) bisimilar arguments (c and y) in two argumentation models.

Definition 7 (Total 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':*

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 $(\mathcal{M}, a) \Leftrightarrow (\mathcal{M}', a')$.*

Now, since logic $\mathsf{K}^{\forall}$ is invariant under total bisimulation [1] and logic K^{μ} under bisimulation [15], 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 models, and let Z be a total bisimulation between $\mathcal{M}$ and $\mathcal{M}'$. It holds that:*

$$\begin{aligned}
\mathcal{M}, a \models CFree(\varphi) \wedge \varphi &\iff \mathcal{M}', a' \models CFree(\varphi) \wedge \varphi \\
\mathcal{M}, a \models Adm(\varphi) \wedge \varphi &\iff \mathcal{M}', a' \models Adm(\varphi) \wedge \varphi \\
\mathcal{M}, a \models Complete(\varphi) \wedge \varphi &\iff \mathcal{M}', a' \models Complete(\varphi) \wedge \varphi \\
\mathcal{M}, a \models Stable(\varphi) \wedge \varphi &\iff \mathcal{M}', a' \models Stable(\varphi) \wedge \varphi \\
\mathcal{M}, a \models Grounded &\iff \mathcal{M}', a' \models Grounded.
\end{aligned}$$

Proof. Follows directly from the fact that bisimulation implies K^{μ}-equivalence [15], and total bisimulation implies $\mathsf{K}^{\forall}$-equivalence [1]. □

In other words, Theorem 7 states that if two arguments are totally bisimilar, then they are indistinguishable from the point of view of abstract argumentation in the sense that the first belongs to a given conflict-free, or admissible set φ if and only if also the second does, and the first belongs to a given stable, complete extension φ, or to the grounded extension, if and only if also the second does.

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

Table 5: Rules of the bisimulation game for $\mathsf{K}^{\forall}$

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 bisimilar 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 $\mathsf{K}^{\forall}$)**.** *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 rules given in Table 5.*

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 some 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*

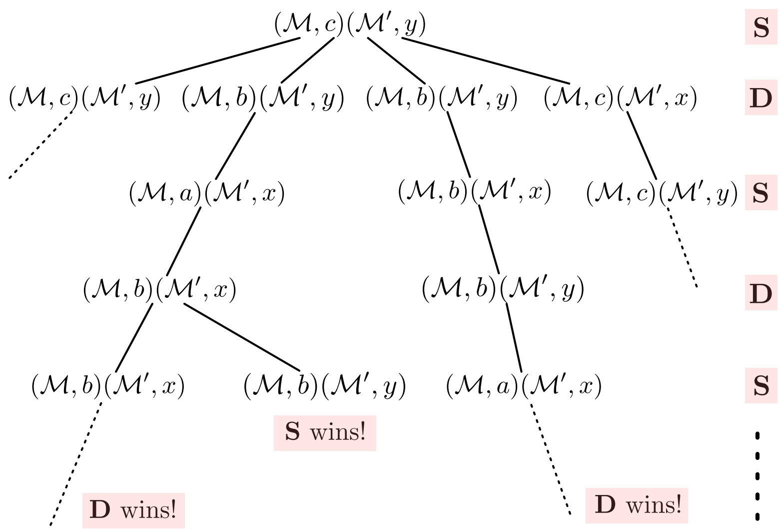

Figure 3: Part of the total bisimulation game played on the models in Figure 2.

$\mathcal{B}(\mathcal{M}, \mathcal{M}')$ *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 have 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:*

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

Proof. The proof is standard and we refer the reader to [15]. □

In other 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)**.** *Let us play a total bisimulation game on the two models* $\mathcal{M}$ *and* $\mathcal{M}'$ *given in 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 3. Notice that* **D** *wins on those infinite paths 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* $((\mathcal{M}, b)(\mathcal{M}', x))$ *by moving in the second model to argument y, which is labelled p while b is not.*

8 Discussion

In this section we address some related work and an important missing piece of our analysis.

8.1 Preferred extensions in modal logic?

The paper has not dealt with one important notion of argumentation: preferred extensions. In [11], preferred extensions are defined as maximal, with respect to set-inclusion, complete extensions. The natural question is whether the logics we have introduced are expressive enough to capture this notion too.

Technically, this means looking for a formula $\varphi(p)$ such that for any pointed model $\mathcal{M} = ((\mathcal{A}, \mathcal{I}), a)$ $\mathcal{M}, a \models \varphi(p)$ iff $a \in ||p||_{\mathcal{M}}$ and $||p||_{\mathcal{M}}$ is a preferred extension of $\mathcal{A}$, where $p \in \mathbf{P}$. It is easy to see that such $\varphi(p)$ can be expressed in monadic second-order logic with a Π^1_1 quantification:

$$p \wedge ST_x(Compl(p)) \wedge \forall q(ST_x(Compl(q)) \rightarrow \neg(p \sqsubset q)) \quad (26)$$

where $ST_x(Compl(p))$ denotes the standard translation [1] of the $\mathsf{K}^{\forall}$ formula for complete extensions (Formula 17) and $q \sqsubseteq p$ means just that $||q||_{\mathcal{M}} \subseteq ||p||_{\mathcal{M}}$, i.e., the truth set of q is included in the truth-set of p. Now the good news is that Formula 26 turns out to be invariant under total bisimulation (Definition 7).

Theorem 9 (Preferred and total bisimulation)**.** *Take $\varphi(p)$ to be defined as in Formula 26 and let $\rightleftharpoons$ denote a total bisimulation relation. For any two pointed models $(\mathcal{M}, a)$ and $(\mathcal{M}', a')$ it holds that:*

$$(\mathcal{M}, a) \rightleftharpoons (\mathcal{M}', a') \implies (\mathcal{M}, a) \models \varphi(p) \iff \mathcal{M}', a' \models \varphi(p)$$

(Sketch). Assume per absurdum that $\mathcal{M}', a' \models \exists q(ST_x(Compl(q)) \wedge (p \sqsubset q))$. By Definition 7 and Theorem 7 we obtain $\mathcal{M}, a \models q \wedge ST_x(Compl(q)) \wedge (p \sqsubset q)$ which contradicts the assumption. The other direction is similar. □

In short, Theorem 9 states that the monadic second-order formula expressing preferred extensions is invariant under total bisimulation. So, although not expressible in K^{μ}, which is precisely equivalent to the bisimulation invariant fragment of monadic second-order [26], Formula 26 should be expressible in K^{μ} extended with the universal modality. Such formulation, which should rely on a smart use of the μ operator, still defies us and is left for future work. Notice also that as a consequence of Theorem 9, Theorem 7 carries over to preferred extensions.

8.2 Related work

To the best of our knowledge, only two papers have dealt with the relationship between logic and argumentation theory. The first one is [3] which presents preliminary work aimed at generalizing abstract argumentation within a logical language. There are two main differences with our approach: first, propositional atoms denote arguments instead of sets of arguments; second, the various extensions, instead of being defined in the logic, are taken to be primitives. The resulting logic is non-standard and no proof procedures (e.g., calculi or games) nor meta-theoretical results are studied.

The second one [7] is closer in purpose to our work. It aims at defining several notions of extensions within modal logic. However, while our approach is eminently model-theoretical, [7] proceeds from a proof-theoretic point of view, characterizing complete and grounded extensions within provability logic. Unlike in our approach, also [7] uses propositional atoms to denote arguments rather than sets thereof.

9 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:

- Find a K^{μ} formula (possibly extended with universal modality) expressing preferred extensions.
- Apply the same methods to obtain modal-logic formulations of other argumentation-theoretic notions, such as semi-stable sets[6].
- Investigate MSO model-checking games as a more general logical setting for dialogue games than the modal model-checking games presented in the paper.
- Develop a systematic comparison of model-checking games and standard dialogue games for argumentation.
- 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 [10].
- Apply sabotage modal logic [24] to study the robustness of the membership of an argument to a certain set or extension denoted by a formula φ.
- Apply the methods and techniques developed in dynamic logic [25] for the "dynamification" of modal logic to study the dynamics of argumentation.

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 sparked 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] G. Boella, J. Hulstijn, and L. van der Torre. A logic of abstract argumentation. In S. Parsons, N. Maudet, P. Moraitis, and Y. Rahwan, editors, *Proceedings of ArgMAS 2005*, pages 29–41. Springer, 2006.

[4] 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.

[5] 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.

[6] 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.

[7] M. Caminada and D. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2):109–145, 2009.

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

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

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

[11] 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.

[12] P. Dunne and T. Bench-Capon. Complexity and combinatorial properties of argument systems. Technical report, University of Liverpool, 2001.

[13] 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.

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

[15] 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.

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

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

[18] 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.

[19] E. Oikarinen and S. Woltran. Characterizing strong equivalence for argumentation frameworks. In F. Lin, U. Sattler, and M. Truszczynski, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference (KR 2010)*, Toronto, Ontario, Canada, May 9–13 2010. AAAI Press.

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

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

[22] 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.

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

[24] J. van Benthem. An essay on sabotage and obstruction. In D. Hutter and W. Stephan, editors, *Mechanizing Mathematical Reasoning*, volume 2605 of *LNCS*, pages 268–276. Springer, 2005.

[25] H. van Ditmarsch, B. Kooi, and W. van der Hoek. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library Series*. Springer, 2007.

[26] Y. Venema. Lectures on the modal μ-calculus. Renmin University in Beijing (China), 2008.

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

[28] 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.

A 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 A briefly recapitulates the key notions developed in [11] which are considered in the paper. For an explanation of the order-theoretic notions involved in the definitions we refer the reader to [9].

The notions in Table A 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 $c_{\mathcal{A}}(X)$. A set X is said to be acceptable with respect to a set Y if and only if all its arguments are defended by arguments in Y. The notion of conflict-freeness is self-explanatory. An admissible set is a set of arguments X which is condlict-free and is able to defend all its attackers. So, admissible sets can be thought of as 'admissible' positions within an argumentation. By considering those admissible sets which contain all their defenders, we obtain the notion of complete extension, which somehow formalizes the idea of a fully exploited admissible position, 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

$c_{\mathcal{A}}$ 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 is acceptable w.r.t. Y in $\mathcal{A}$	iff	$X \subseteq c_{\mathcal{A}}(Y)$
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 is conflict-free and $X \subseteq c_{\mathcal{A}}(X)$
	iff	X is a conflict-free post-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 stable extension of $\mathcal{A}$	iff	X is a complete extension of $\mathcal{A}$
		and $\forall b \notin X, \exists a \in X : a \rightarrow b$
	iff	$X = \{a \in A \mid \nexists b \in X : b \rightarrow a\}$
X grounded extension of $\mathcal{A}$	iff	X is the minimal complete extension of $\mathcal{A}$
	iff	X is the least fixpoint of $c_{\mathcal{A}}$
X preferred extension of $\mathcal{A}$	iff	X is a maximal complete extension of $\mathcal{A}$

Table 6: Basic notions of argumentation theory.

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.

B Completeness of logic K^{-1}

Theorem 10 (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. □

C A formal proof of the *Fundamental Lemma*

1.	$\varphi \rightarrow \varphi \vee \psi$	**Prop**
2.	$\langle\leftarrow\rangle\varphi \rightarrow \langle\leftarrow\rangle(\varphi \vee \psi)$	1, K – derived rule
3.	$[\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle(\varphi \vee \psi)$	2, K – derived rule
4.	$(\alpha \vee \beta \rightarrow \gamma) \rightarrow (\beta \rightarrow \gamma)$	**Prop**
5.	$(\psi \vee \xi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \rightarrow (\xi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	4, instance
6.	$(\psi \vee \xi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \rightarrow (\xi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi \vee \psi)$	5, 3, **Prop**, **MP**
7.	$[\forall](\psi \vee \xi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \rightarrow [\forall](\xi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi \vee \psi)$	6, K – derived rule
8.	$Acc(\psi \vee \xi, \varphi) \rightarrow Acc(\xi, \varphi \vee \psi)$	7, definition
9.	$(\psi \vee \xi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \rightarrow (\psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	4, instance
10.	$[\forall](\psi \vee \xi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \rightarrow [\forall](\psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	9, K – derived rule
11.	$Acc(\psi \vee \xi, \varphi) \rightarrow Acc(\psi, \varphi)$	10, definition
12.	$((\alpha \rightarrow \gamma) \wedge (\beta \rightarrow \gamma)) \rightarrow (\alpha \vee \beta \rightarrow \gamma)$	**Prop**
13.	$([\forall](\alpha \rightarrow \gamma) \wedge [\forall](\beta \rightarrow \gamma)) \rightarrow [\forall](\alpha \vee \beta \rightarrow \gamma)$	12, **N**, **K**, **MP**
14.	$([\forall](\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \wedge [\forall](\psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)) \rightarrow [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	13, Instance
15.	$[\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle(\varphi \vee \psi)$	14, **Prop**, **K**, **N**
16.	$([\forall](\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \wedge [\forall](\psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)) \rightarrow [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi \vee \psi)$	15, **Prop**, **K**, **N**
17.	$Acc(\varphi, \varphi) \wedge Acc(\psi, \varphi) \rightarrow Acc(\varphi \vee \psi, \varphi \vee \psi)$	16, definition
18.	$Acc(\varphi, \varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow Acc(\varphi \vee \psi, \varphi \vee \psi)$	17, 9, **Prop**, **MP**
19.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi) \rightarrow [\leftarrow](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi)$	**Incl**
20.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi) \rightarrow ([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\neg\varphi)$	19, **Prop**, **MP**
21.	$[\forall][\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi) \rightarrow [\forall]([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\neg\varphi)$	20, K – derived rule

22.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi) \rightarrow [\forall]([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\neg\varphi)$	21, S5 – derived rule
23.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi) \wedge [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	
	$\rightarrow [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \wedge [\forall]([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\neg\varphi)$	22, **Prop**, **MP**
24.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi) \wedge [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \rightarrow [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\neg\varphi)$	23, **Prop**, **MP**
25.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi \wedge \neg\psi) \rightarrow [\leftarrow](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi \wedge \neg\psi)$	**Incl**
26.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi \wedge \neg\psi) \rightarrow ([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\neg\varphi \wedge \neg\psi)$	25, **K**, **Prop**, **MP**
27.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi \wedge \neg\psi) \rightarrow [\forall]([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\neg\varphi \wedge \neg\psi)$	26, S5 – derived rule
28.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi) \wedge [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	
	$\rightarrow [\forall]([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\neg\varphi \wedge \neg\psi)$	24, 27, **Prop**, **MP**
29.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi) \wedge [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	
	$\rightarrow [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \wedge [\forall]([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\neg\varphi \wedge \neg\psi)$	28, **Prop**, **MP**
30.	$[\forall](\alpha \rightarrow \beta) \wedge [\forall](\beta \rightarrow \gamma) \rightarrow [\forall](\alpha \rightarrow \gamma)$	S5 – theorem
31.	$[\forall]([\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow](\neg\varphi \wedge \neg\psi)) \wedge [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	
	$\rightarrow [\forall](\varphi \vee \psi \rightarrow [\leftarrow](\neg\varphi \wedge \neg\psi))$	30, instance
32.	$[\forall](\langle\leftarrow\rangle\varphi \rightarrow \neg\varphi) \wedge [\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	
	$\rightarrow [\forall](\varphi \vee \psi \rightarrow [\leftarrow](\neg\varphi \wedge \neg\psi))$	29, 31, **Prop**, **MP**
33.	$CFree(\varphi) \wedge Acc(\varphi \vee \psi, \varphi) \rightarrow CFree(\varphi \vee \psi)$	32, definition
34.	$Acc(\varphi, \varphi) \wedge Acc(\psi, \varphi) \rightarrow Acc(\varphi \vee \psi, \varphi)$	14, definition
35.	$CFree(\varphi) \wedge Acc(\varphi, \varphi) \wedge Acc(\psi, \varphi) \rightarrow CFree(\varphi \vee \psi)$	33, 34, **Prop**, **MP**
36.	$CFree(\varphi) \wedge Acc(\varphi, \varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow CFree(\varphi \vee \psi)$	35, 9, **Prop**, **MP**
37.	$CFree(\varphi) \wedge Acc(\varphi, \varphi) \wedge Acc(\psi \vee \xi, \varphi)$	
	$\rightarrow CFree(\varphi \vee \psi) \wedge Acc(\varphi \vee \psi, \varphi \vee \psi)$	36, 18, **Prop**, **MP**
38.	$CFree(\varphi) \wedge Acc(\varphi, \varphi) \wedge Acc(\psi \vee \xi, \varphi)$	
	$\rightarrow CFree(\varphi \vee \psi) \wedge Acc(\varphi \vee \psi, \varphi \vee \psi) \wedge Acc(\xi, \varphi \vee \psi)$	37, 8, **Prop**, **MP**
39.	$Adm(\varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow Adm(\varphi \vee \psi) Acc(\xi, \varphi \vee \psi)$	38, definition