On the Logic of Argumentation Theory

Davide Grossi
Institute for Logic, Language and Computation
University of Amsterdam
Science Park 904, 1098 XH
Amsterdam, The Netherlands
d.grossi@uva.nl

ABSTRACT

The paper applies modal logic to formalize fragments of argumentation theory. Such formalization allows to import, for free, a wealth of new notions (e.g., argument equivalence), new techniques (e.g., calculi, model-checking games, bisimulation games), and results (e.g., completeness of calculi, adequacy of games, complexity of model-checking) from logic to argumentation.

Categories and Subject Descriptors

I.2.4 [**Knowledge Representation Formalisms and Methods**]: Modal logic

General Terms

Theory

Keywords

Argumentation theory, modal logic

1. INTRODUCTION

The paper analyzes argumentation from the point of view of formal logic. It shows how standard results in argumentation theory obtain elegant reformulations within well-investigated modal logics. This allows to import—for free—a number of techniques (e.g., calculi, logical games) as well as results (e.g. completeness, adequacy, complexity) from modal logic to argumentation theory. Also, as is often the case in the cross-fertilization of different formalisms, this perspective opens up new lines of research which were thus far hidden to the attention of argumentation theorists.

Although the results presented are theoretical, they set the stage for the development of logic-based techniques for argumentation in multi-agent systems such as, eminently, the formal verification (via model-checking) of argumentation systems, and the design of multi-agent argumentation protocols via logic games.

Let us start with the basic structure of argumentation theory. An abstract argumentation framework is a relational structure $\mathcal{A} = (A, \rightarrow)$ where A is a non-empty set of arguments, and $\rightarrow \subseteq A^2$ is an 'attack' relation on A [6]. So, the intuitive reading of $a \rightarrow b$ is that argument a attacks argument b. This paper investigates the simple but yet unexplored idea which consists in viewing abstract argumentation frameworks as Kripke frames (S, R) [1] where $S = A$, that is, the set of states is the set of arguments, and $R = \rightarrow^{-1}$, that is, the accessibility relation is the inverse of the attack relation or, intuitively, the 'being attacked' relation. The entire paper hinges on this simple assumption.

For space reasons the paper cannot introduce argumentation theory in an extensive way but, to make it as most self-contained as possible, the main argumentation-theoretic notions from [6] have been recapitulated in Table 1. As such notions are formalized along the paper, their intuitive reading will also be provided. This said, the paper is organized as follows. Section 2 starts off by applying a well-known modal logic to study a first set of notions of argumentation theory. This enables the possibility of using calculi to derive argumentation-theoretic results such as the Fundamental Lemma [6], and import complexity results concerning, for instance, checking whether a given set is a stable extension. Along the same line, Section 3 tackles the formalization of the notion of grounded extension within the modal μ-calculus. In Section 4 semantic games are studied for the logic introduced in Section 2 which provide a version of dialogue games as model-checking games. Section 5 tackles the question—not yet addressed in the literature—of when two arguments, or two argumentation frameworks, are "indistinguishable" from the point of view of argumentation theory. For this purpose the model-theoretic notion of bisimulation is introduced and bisimulation games are presented as a procedural method to check the "behavioral equivalence" of two argumentation frameworks. Section 6 addresses the problem of the representation of preferred extensions, briefly discusses related work and concludes.

Cite as: On the Logic of Argumentation Theory, Davide Grossi, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 409-416

2. ARGUMENTS IN MODAL DISGUISE

2.1 Argumentation models

If an argumentation framework can be viewed as a Kripke frame, 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*

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

Table 1: Basic notions of argumentation theory (X denotes a set of arguments).

$\mathfrak{A}$. *A pointed argumentation model is a pair $(\mathcal{M}, a)$ where $\mathcal{M}$ is an argumentation model and a an argument from A.*

Argumentation models are nothing but argumentation frames together with a way of "naming" sets of arguments or, to put it otherwise, of "labeling" arguments. The fact that an argument a belongs to $\mathcal{I}(p)$ in a given model $\mathcal{M}$, which in logical notation reads $(\mathcal{A}, \mathcal{I}), a \models p$, can be interpreted as stating that "argument a has property p" , or that "p is true of a". By substituting p with a Boolean compound φ (e.g., $\varphi := p \wedge q$) 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 [3].

EXAMPLE 1. *(Argument labelings as argumentation models) In argumentation theory, a labeling function [3] is a function $l : \{\mathtt{1}, \mathtt{0}, ?\} \longrightarrow A$ from the set of three labels $\{$ `1`, `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* $\mathtt{Label} := (\mathtt{1} \wedge \neg\mathtt{0} \wedge \neg ?) \vee (\neg\mathtt{1} \wedge \mathtt{0} \wedge \neg ?) \vee (\neg\mathtt{1} \wedge \neg\mathtt{0} \wedge ?)$. *Hence, a framework $\mathcal{A}$ with a labeling function is nothing but an argumentation model $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ s.t. $\mathcal{M} \models \mathtt{Label}$.*

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".

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

This section introduces logic $\mathsf{K}^\forall$, an extension of the minimal modal logic K with universal modality.

2.2.1 Language.

The language of $\mathsf{K}^\forall$ is a standard modal language with two modalities: $\langle\leftarrow\rangle$ and $\langle\forall\rangle$, i.e., the universal modality. It is 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 $\{\top, \vee, \rightarrow\}$ and modal $\{[\leftarrow], [\forall]\}$ connectives are defined as usual.

2.2.2 Semantics.

DEFINITION 2 (SATISFACTION). *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):*

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

As usual, 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., $\mathfrak{M} \models \varphi$. The truth-set of a formula φ is denoted $|\varphi|_{\mathcal{M}}$.

Logic $\mathsf{K}^\forall$ is therefore endowed with modal operators of the type "there exists an argument attacking the current one such that", i.e., $\langle\leftarrow\rangle$, and "there exists an argument such that", i.e., $\langle\forall\rangle$, together with their duals. Given an argumentation model $\mathcal{M}$ we can thereby express statements such as the ones adverted to above: "a is attacked by an argument in a set called φ" corresponds to $\langle\leftarrow\rangle\varphi$ being true in the pointed model $(\mathcal{M}, a)$ and "a is defended by the set φ" corresponds to $\langle\leftarrow\rangle\langle\leftarrow\rangle\varphi$ being true in the pointed model $(\mathcal{M}, a)$.

2.2.3 Axiomatics.

Logic $\mathsf{K}^\forall$ is axiomatized as follows, where $i \in \{\leftarrow, \forall\}$:

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

2.2.4 Meta-theoretical results.

We list the following known relevant results.

- Logic $\mathsf{K}^\forall$ is sound and strongly complete for the class $\mathfrak{A}$ of argumentation frames [1, Ch. 7].
- The complexity of checking whether a formula of $\mathcal{L}^{\mathsf{K}^\forall}$ is satisfied by a pointed model $\mathcal{M}$ is P-complete [10].

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

Perhaps surprisingly, logic $\mathsf{K}^\forall$ is already expressive enough to capture several basic notions of argumentation theory

such as: conflict freeness, acceptability, admissibility, complete extensions, stable extensions.

$$Acc(\varphi,\psi) := [\forall](\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\psi) \quad (1)$$
$$CFree(\varphi) := [\forall](\varphi \rightarrow [\leftarrow]\neg\varphi) \quad (2)$$
$$Adm(\varphi) := [\forall](\varphi \rightarrow ([\leftarrow]\neg\varphi \wedge [\leftarrow]\langle\leftarrow\rangle\varphi)) \quad (3)$$
$$Compl(\varphi) := [\forall]((\varphi \rightarrow [\leftarrow]\neg\varphi) \wedge (\varphi \leftrightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)) \quad (4)$$
$$Stable(\varphi) := [\forall](\varphi \leftrightarrow [\leftarrow]\neg\varphi) \quad (5)$$

Intuitively, 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 1). 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 2). A set of arguments φ is admissible if and only if it is conflict free and acceptable with respect to itself (Formula 3). 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 4). 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 5). The adequacy of these definitions with respect to the standard ones in Table 1 is easily checked.

The following examples applies $\mathsf{K}^\forall$ to Example 1.

EXAMPLE 2. *(Argumentation labelings in $\mathsf{K}^\forall$) According to [3], 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:*

$$[\forall]((\mathtt{1} \leftrightarrow [\leftarrow]\mathtt{0}) \wedge (\mathtt{0} \leftrightarrow \langle\leftarrow\rangle\mathtt{1}) \wedge \mathtt{Label}) \quad (6)$$

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 6. Also, it is a matter of propositional reasoning to see that Formula 6 is equivalent to the following formula:*

$$Compl(\mathtt{1}) \wedge [\forall]((\mathtt{0} \leftrightarrow \langle\leftarrow\rangle\mathtt{1}) \wedge \mathtt{Label}) \quad (7)$$

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 4) 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.*

We can now prove results of argumentation theory, such as the ones proven in [6], which concern the notions formalized in Formulae 1-5 as theorems of $\mathsf{K}^\forall$.

THEOREM 1 ([6] FORMALIZED). *The following formulae are theorems of $\mathsf{K}^\forall$:*

$$Adm(\varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow Adm(\varphi \vee \psi) \wedge Acc(\xi, \varphi \vee \psi) \quad (8)$$
$$Stable(\varphi) \rightarrow Adm(\varphi) \quad (9)$$
$$Stable(\varphi) \rightarrow Compl(\varphi) \quad (10)$$

PROOF (SKETCH). The theorem is easily proven semantically by then calling in completeness. However, as an example of the application of the calculus, we provide in Figure 1 the derivation of a sub-result of Formula 8: The proof is completed by proving that: $Adm(\varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow Acc(\varphi,\varphi) \wedge Acc(\psi,\varphi)$, $Adm(\varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow CFree(\varphi \vee \psi)$, and $Adm(\varphi) \wedge Acc(\psi \vee \xi, \varphi) \rightarrow Acc(\xi, \varphi \vee \psi)$. □

$((\alpha \rightarrow \gamma) \wedge (\beta \rightarrow \gamma)) \rightarrow (\alpha \vee \beta \rightarrow \gamma)$	**Prop**
$([\forall](\alpha \rightarrow \gamma) \wedge [\forall](\beta \rightarrow \gamma)) \rightarrow [\forall](\alpha \vee \beta \rightarrow \gamma)$	$2, \mathbf{N}, \mathbf{K}, \mathbf{MP}$
$([\forall](\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \wedge [\forall](\psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)) \rightarrow$	
$[\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)$	Instance of 3
$[\leftarrow]\langle\leftarrow\rangle\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle(\varphi \vee \psi)$	$\mathbf{Prop}, \mathbf{K}, \mathbf{N}$
$([\forall](\varphi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi) \wedge [\forall](\psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi)) \rightarrow$	
$[\forall](\varphi \vee \psi \rightarrow [\leftarrow]\langle\leftarrow\rangle\varphi \vee \psi)$	$4, \mathbf{Prop}, \mathbf{K}, \mathbf{N}$
$Acc(\varphi,\varphi) \wedge Acc(\psi,\varphi) \rightarrow Acc(\varphi \vee \psi, \varphi \vee \psi)$	5, definition

Figure 1: Example of a derivation in $\mathsf{K}^\forall$.

Formula 8 is a generalized version of the so-called *Fundamental Lemma* proven in [6]. 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 $\varphi \vee \psi$. Formulae 9 and 10 state well-known facts about the relative strength of admissible, complete and stable extensions.

Other results can be formalized along the same lines. What the section has shown 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. Besides, by the results reported in Section 2.2.4 it follows that model-checking whether a given formula is conflict free, admissible, acceptable (with respect to another formula), complete or stable can be done in polynomial time: e.g., "$\mathcal{M}, a \models Stable(\varphi)$?". Similarly, it follows that it can be checked in polynomial time whether an argument belongs to the truth-set of a formula which is conflict free, admissible, acceptable (with respect to another formula), complete or stable: e.g., "$\mathcal{M}, a \models \varphi \wedge Stable(\varphi)$?".

3. MODAL 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 [6], the grounded extension is defined as the smallest fixpoint of the characteristic function of an argumentation framework (see Table 1).

3.1 Characteristic functions in K

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., $\{a \in A \mid \forall b \in A : [b \rightarrow a \Rightarrow \exists c \in X : c \rightarrow b]\}$. Does logic $\mathsf{K}^\forall$ 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 $\mathbf{P}$. Language $\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}$ is the fragment of $\mathcal{L}^{\mathsf{K}^\forall}$ containing only the compounded modal operator $[\leftarrow]\langle\leftarrow\rangle$ or, also, simply the fragment of $\mathcal{L}^{\mathsf{K}}$ (i.e., f $\mathcal{L}^{\mathsf{K}^\forall}$ without universal modality) containing only the $[\leftarrow]\langle\leftarrow\rangle$-operator. Let $\mathcal{A}^+ = (2^A, \cap, -, \emptyset, c_\mathcal{A})$ be the power set alge-

bra on $2^{\mathcal{A}}$ 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 $\mathcal{I}^* : \mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle} \longrightarrow 2^{\mathcal{A}}$ be the inductive extension of a valuation function $\mathcal{I} : \mathbf{P} \longrightarrow 2^{\mathcal{A}}$ according to the semantics given in Definition 2. We can prove the following result.

THEOREM 2 ($c_{\mathcal{A}}$ VS. $[\leftarrow]\langle\leftarrow\rangle$). *Let $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ be an argumentation model. Function $\mathcal{I}^*$ is a homomorphism from $\mathfrak{ter}_{\mathcal{L}^{[\leftarrow]\langle\leftarrow\rangle}}$ to $\mathcal{A}^+$.*

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

$$\begin{aligned} |[\leftarrow]\langle\leftarrow\rangle\varphi|_{\mathcal{M}} &= \{a \mid \forall b : a \rightarrow^{-1} b, \exists c : b \rightarrow^{-1} c \;\&\; c \in |\varphi|_{\mathcal{M}}\} \\ &= \{a \mid \forall b : b \rightarrow a, \exists c : c \rightarrow b \;\&\; c \in |\varphi|_{\mathcal{M}}\} \\ &= c_{\mathcal{A}}(|\varphi|_{\mathcal{M}}). \end{aligned}$$

This completes the proof. □

In other words, Theorem 2 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 φ of arguments. Notice also that from Theorem 2 the adequacy of Formulae 1-5 with respect to the definitions in Table 1 follows straightforwardly.

Characteristic functions are known to be monotonic [6] hence, by Theorem 2, we get that $[\leftarrow]\langle\leftarrow\rangle$ denotes a monotonic function and therefore, by the Knaster-Tarski theorem[1] we have that there always exist a greatest and a least $[\leftarrow]\langle\leftarrow\rangle$-fixpoint. From a logical point of view this means that, in order to be able to express the grounded extension, it suffices to add to the K fragment of $\mathsf{K}^{\forall}$ a least fixpoint operator. This takes us to the realm of μ-calculi.

3.2 μ-calculus for argumentation

3.2.1 Language.

To add the least fixpoint operator μ to logic K we first define language $\mathcal{L}^{\mathsf{K}^{\mu}}$ via the following BNF:

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

where p ranges over $\mathbf{P}$ and $\varphi(p)$ indicates that p occurs free in φ (i.e., it is not bounded by fixpoint operators) and under an even number of negations.[2] 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 $\mathcal{L}^{\mathsf{K}^{\mu}}$.

3.2.2 Semantics.

DEFINITION 3 (SATISFACTION). *Let $\varphi \in \mathcal{L}^{\mathsf{K}^{\mu}}$. The satisfaction of φ by a pointed model $(\mathcal{M}, a)$, with $\mathcal{M} = (\mathcal{A}, \mathcal{I})$, is inductively defined as follows (Boolean clauses, as well as the clause for $\langle\leftarrow\rangle$, are as in Definition 2):*

$$\mathcal{M}, a \models \mu p.\varphi(p) \quad \textit{iff} \quad a \in \bigcap\{X \in 2^{\mathcal{A}} \mid |\varphi|_{\mathcal{M}[p:=X]} \subseteq X\}$$

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., $\mathfrak{M} \models \varphi$.

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:

$$\textit{Grounded} := \mu p.[\leftarrow]\langle\leftarrow\rangle p \tag{11}$$

Notice that, unlike the notions formalized in Formulae 1-5, the grounded extension of a framework is always unique and does not depend on the particular labeling of a given model.

3.2.3 Axiomatics.

Logic K^{μ} is axiomatized by the following rules and axiom schemata.

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

3.2.4 Meta-theoretical results.

We list two 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 3 [14]. Notice however that, unlike $\mathsf{K}^{\forall}$, the given axiomatics of K^{μ} is not strongly complete since it is obviously not compact.
- The complexity of the model-checking problem for a formula of size m and alternation depth d on a system of size n is $O(m \cdot n^{d+1})$ [8] where the alternation depth of a formula of $\mathcal{L}^{\mathsf{K}^{\mu}}$ is the maximum number of $\mu/\neg\mu\neg$ in a chain of nested fixpoints.

3.3 Doing argumentation in K^{μ}

Like in Section 2.3 we give a couple of examples of the kind of argumentation-theoretic results formalizable in K^{μ}.

THEOREM 3 (GROUNDED EXTENSION IS CONFLICT-FREE). *The following formula is a validity of K^{μ}:*

$$\textit{Grounded} \rightarrow \neg[\leftarrow]\textit{Grounded} \tag{12}$$

PROOF. We proceed per absurdum applying the definition in Formula 11. Take an argumentation model satisfying Formula 12 and assume that there exist arguments a, b such that $a \rightarrow^{-1} b$ and $\mathcal{M}, b \models \mu p.[\leftarrow]\langle\leftarrow\rangle p$ while also

[1] We refer the interested reader to [5].

[2] This syntactic restriction guarantees that every formula $\varphi(p)$ defines a monotonic set transformation.

$\mathcal{M}, a \models \mu p.[\leftarrow]\langle\leftarrow\rangle p$. We distinguish two cases: 1) there exists a finite chain ($a \rightarrow^{-1} b \rightarrow^{-1} b_1 \rightarrow^{-1} \ldots \rightarrow^{-1} b_n$) 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 3, 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: $\{c \in A \mid a \rightarrow_{2m}^{-1} c\}$ and $\{c \in A \mid b \rightarrow_{2m}^{-1} c\}$ where $\rightarrow_{2m}^{-1}$ denotes reachability via $\rightarrow^{-1}$ in an even number of steps. We thus obtain a contradiction. □

Like Theorem 2, Theorem 3 provides a modal logic formulation of an argumentation-theoretic result. Let us now look at the complexity.

THEOREM 4 (MODEL-CHECKING GROUNDED). *Let $\mathcal{M}$ be an argumentation model. It can be decided in polynomial time whether an argument a belongs to the grounded extension of $\mathcal{M}$, that is, whether $\mathcal{M}, a \models$ Grounded.*

PROOF. Since $\mu p.[\leftarrow]\langle\leftarrow\rangle p$ has alternation depth 0 it follows, by the result reported in Section 3.2.4, 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 $\mathcal{M}$. □

4. DIALOGUE GAMES & LOGIC GAMES

The proof-theory of abstract argumentation is commonly given in terms of dialogue games [12]. The present section introduces a new game-theoretic proof procedure for argumentation theory based on model-checking games. In model-checking games, 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 deals with the model-checking game for $\mathsf{K}^{\forall}$. For K^{μ}-variant we refer the reader to [13].

4.1 Model-checking game for $\mathsf{K}^{\forall}$

A model-checking 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.

DEFINITION 4 ($\mathsf{K}^{\forall}$-MODEL-CHECKING GAME). *Let $\varphi \in \mathcal{L}^{\mathsf{K}^{\forall}}$, and $\mathcal{M}$ be an argumentation model. The model-checking game $\mathcal{C}(\varphi, \mathcal{M})$ is defined by the following items.* **Players:** *The set of players is $\{\exists, \forall\}$. An element from $\{\exists, \forall\}$ will be denoted P and its opponent $\overline{P}$.* **Game form:** *The* game form *of $\mathcal{C}(\varphi, \mathcal{M})$ is defined by the board game in Table 2.* **Winning conditions:** *Player P wins if and only if $\overline{P}$ has to play in a position with no available moves.* **Instantiation:** *The* instance *of $\mathcal{C}(\varphi, \mathcal{M})$ with starting point (φ, a) is denoted $\mathcal{C}(\varphi, \mathcal{M})@(\varphi, a)$.*

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 2: Rules of the model-checking game for $\mathsf{K}^{\forall}$.

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: ∃ if the formula is a disjunction, a box, a false atom or ⊥, and ∀ in the remaining cases.[3]

DEFINITION 5 (WINNING STRATEGIES AND POSITIONS). *A strategy for player P in $\mathcal{C}(\varphi, \mathcal{M})@(\varphi, a)$ is a function telling P what to do in any match played from position (φ, a). Such a strategy is* winning for *P if and only if, in any match played according to the strategy, P wins. A position (φ, a) in $\mathcal{C}(\varphi, \mathcal{M})$ is* winning for *P if and only if P has a winning strategy in $\mathcal{C}(\varphi, \mathcal{M})@(\varphi, a)$. The set of winning positions of $\mathcal{C}(\varphi, \mathcal{M})$ is denoted $Win_P(\mathcal{C}(\varphi, \mathcal{M}))$.*

By Definitions 4 and 5 it follows that the model-checking game is a two-players zero-sum game. It is known that such games are determined, that is, each match has a winner [15].

It remains to be proven that the game is adequate with respect to the semantics of $\mathsf{K}^{\forall}$. To put it otherwise, we have to prove that if ∃ 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 ∃ always wins the corresponding game instantiated at that point.

THEOREM 5 (ADEQUACY). *Let $\varphi \in \mathcal{L}^{\mathsf{K}^{\forall}}$, and let $\mathcal{M} = (\mathcal{A}, \mathcal{I})$ be an argumentation model. Then, for all $a \in A$:*

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

PROOF (SKETCH). The proof is by induction on the length l of φ. A proof for K without the universal modality can be found in [13]. It suffices to extend the inductive case to cover formulae with the universal modality. The base case $l = 0$ is straightforward. For the step $l > 0$ we provide a proof of the modal case $\varphi = \langle\forall\rangle\psi$. From left to right. Assume $(\varphi, a) \in Win_{\exists}(\mathcal{C}(\varphi, \mathcal{M}))$. It is ∃'s turn to move. It follows that there exists a position (ψ, b) s.t. it is a winning position for ∃. By induction hypothesis we conclude that $\mathcal{M}, b \models \psi$ and hence $\mathcal{M}, a \models \langle\forall\rangle\psi$. From right to left. Assume $\mathcal{M}, a \models \varphi$. It follows that there exists b s.t. $\mathcal{M}, b \models \psi$. By induction hypothesis we have that $(\psi, b) \in Win_{\exists}(\mathcal{C}(\psi, \mathcal{M}))$. But it is ∃'s turn to move, hence we conclude $(\varphi, a) \in Win_{\exists}(\mathcal{C}(\varphi, \mathcal{M}))$. □

4.2 Games for model-checking extensions

The next example illustrates a model-checking game for stable extensions run on the so-called Nixon diamond [12].

[3] Notice that positions use formulae in positive normal form.

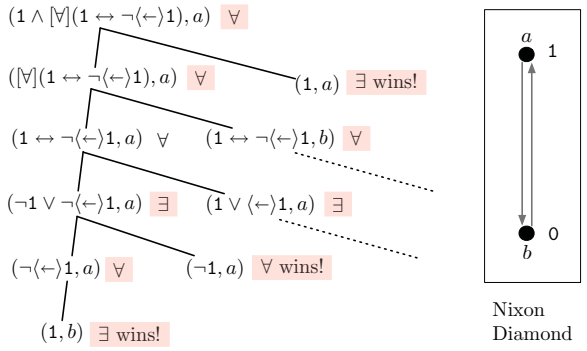

Figure 2: Game for stable extensions in the 2-cycle.

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 2). 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 $\mathcal{C}(\mathtt{1} \wedge \mathit{Stable}(\mathtt{1}), (\mathcal{A}, \mathcal{I}))$ initialized at position $(\mathtt{1} \wedge \mathit{Stable}(\mathtt{1}), a)$. That is, spelling out the definition of $\mathit{Stable}(\mathtt{1})$: $\mathcal{C}(\mathtt{1} \wedge [\forall](\mathtt{1} \leftrightarrow \neg\langle\leftarrow\rangle\mathtt{1}))@(\mathtt{1} \wedge [\forall](\mathtt{1} \leftrightarrow \neg\langle\leftarrow\rangle\mathtt{1}), a)$. Such a game, played according to the rules in Definitions 4 and 5, gives rise to the tree in Figure 2.*

In general, model-checking games provide a proof procedure for checking whether an argument belongs to a certain extension given an argumentation model, which we have seen in Sections 2.3 and 3.3 to be a polynomial problem. The structure of such proof procedure is invariant, and the different games are obtained simply by changing the formula to be checked (Table 3).[4] This feature confers a high systematic flavor to this sort of games for argumentation.

Now the natural question arises of what the precise relationship is between model-checking games and the sort of games studied in argumentation, called dialogue games [12].

4.3 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 2.3 and 3.3, 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 [7]). 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.[5]

[4]Note that the game for checking grounded extensions is, obviously, the model-checking game for K^μ [13].

Admis. :	$\mathcal{C}(\varphi \wedge \mathit{Adm}(\varphi), \mathcal{M})@(\varphi \wedge \mathit{Adm}(\varphi), a)$
Compl. :	$\mathcal{C}(\varphi \wedge \mathit{Compl}(\varphi), \mathcal{M})@(\varphi \wedge \mathit{Compl}(\varphi), a)$
Stable :	$\mathcal{C}(\varphi \wedge \mathit{Stable}(\varphi), \mathcal{M})@(\varphi \wedge \mathit{Stable}(\varphi), a)$
Grounded :	$\mathcal{C}(\mathit{Grounded}, \mathcal{M})@(\mathit{Grounded}, a)$

Table 3: Games for model-checking extensions.

To conclude, we might say that the games defined in Section 4.1 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 3—to dialogue games.

5. INDISTINGUISHABLE ARGUMENTS

Since abstract argumentation neglects the internal structure of arguments, the natural question arises of when two arguments can be said to be the same from the point of view of argumentation theory. Studying such notion of "sameness" or "equivalence" of arguments is not just a mathematical diversion. A simple example where this issue appears is in legal reasoning, and in particular within common-law systems. Often, in such systems the so-called principle of *stare decisis* [11] holds. According to such a principle, a judge should rule cases that are "substantially the same" in the same way. Now, an essential aspect of a judicial case is its argumentation framework, so being the same in this respect seems to mean something like exhibiting the "same argumentative structure". In the present section we present a formal study of this simple intuition based on $\mathsf{K}^\forall$ and K^μ.

5.1 Bisimilar arguments

The logical analysis of abstract argumentation enables us directly with a well-investigated formal notion of "behavioral equivalence" between arguments/points in a model: bisimulation [1, 9]. It is well-known that logic K^μ is invariant under bisimulation [13]. In the present section we will focus on the specific notion of bisimulation which is tailored to $\mathsf{K}^\forall$, also called *total bisimulation*.

Definition 6 (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 a, a' s.t. aZa':* **Atom:** *a and a' are propositionally equivalent;* **Zig:** *if $a \rightarrow^{-1} b$ for some $b \in A$, then $a' \rightarrow^{-1} lb'$ for some $b' \in A'$ and bZb';* **Zag:** *if $a' \rightarrow^{-1} b'$ for some $b' \in A$ then $a \rightarrow^{-1} b$ for some $\underline{in}A$ and aZa'. A* total bisimulation *is a bisimulation $Z \subseteq A \times A'$ such that its left projection covers A and its right projection covers A'. When a total bisimulation exists between $\mathcal{M}$ and $\mathcal{M}'$ we write $(\mathcal{M}, a) \leftrightarrows (\mathcal{M}', a')$.*

Now, since logic $\mathsf{K}^\forall$ is invariant under total bisimulation [1] and logic K^μ under bisimulation [9], we obtain a natural notion of "sameness" of arguments, which is weaker than the

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

Position	Available moves
$((\mathcal{M},a)(\mathcal{M}',a'))$	$\{((\mathcal{M},a)(\mathcal{M}',b')) \mid \exists b' \in A' : a' \rightarrow^{-1} b'\}$
	$\cup\{((\mathcal{M},b)(\mathcal{M}',a')) \mid \exists b \in A : a \rightarrow^{-1} b\}$
	$\cup\{((\mathcal{M},a)(\mathcal{M}',b')) \mid \exists b' \in A'\}$
	$\cup\{((\mathcal{M},b)(\mathcal{M}',a')) \mid \exists b \in A\}$

Table 4: Rules of the bisimulation game

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.

THEOREM 6 (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:*

$$\mathcal{M},a \models Adm(\varphi) \wedge \varphi \iff \mathcal{M}',a' \models Adm(\varphi) \wedge \varphi$$

where $Adm(\varphi)$ can be substituted by $CFree(\varphi)$, $Compl(\varphi)$ or $Stable(\varphi)$ and $Adm(\varphi) \wedge \varphi$ can be substituted by $Grounded$.

PROOF. Follows directly from the fact that bisimulation implies K^{μ}-equivalence [9], and total bisimulation implies $\mathsf{K}^{\forall}$-equivalence [1]. □

In other words, Theorem 6 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.

5.2 Total bisimulation games

We can associate a game to Definition 6. Such game checks whether two given pointed models $(\mathcal{M},a)$ and $(\mathcal{M}',a')$ are bisimular or not. The game is played by two players: **S**poiler, which tries to show that the two given pointed models are not bisimilar, and **D**uplicator which pursues the opposite goal. A match is started by **S**, then **D** responds, and so on. If and only if **D** moves to a position where the two pointed models are not propositionally equivalent, or if it cannot move, **S** wins.

DEFINITION 7 (TOTAL BISIMULATION GAME). *Take 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 Table 4.* **Turn function:** *If the round is even* **S** *plays, if it is odd* **D** *plays.* **Winning conditions:** **S** *wins if and only if either D has moved to a position $((\mathcal{M},a)(\mathcal{M}',a'))$ where a and a' do not satisfy the same labels, or* **D** *has no available moves. Otherwise* **D** *wins.* **Instantiation:** *The* instance *of $\mathcal{B}(\mathcal{M},\mathcal{M}')$ with starting position $((\mathcal{M},a)(\mathcal{M}',a'))$ is denoted $\mathcal{B}(\mathcal{M},\mathcal{M}')@(a,a')$.*

So, as we might expect, positions in a (total) bisimulation game 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 7, are won by **D**.

From Definition 7 we obtain the following notions of winning strategies and winning positions.

DEFINITION 8 (WINNING STRATEGIES AND POSITIONS). *A strategy for player P in $\mathcal{B}(\mathcal{M},\mathcal{M}')@(a,a')$ is a function telling P what to do in any match played from position (a,a'). Such a strategy is* winning for *P if and only if, in any match played according to the strategy, P wins. A position $((\mathcal{M},a)(\mathcal{M}',a'))$ in $\mathcal{B}(\mathcal{M},\mathcal{M}')$ is* winning for *P if and only if P has a winning strategy in $\mathcal{B}(\mathcal{M},\mathcal{M}')@(a,a')$. The set of all winning positions of game $\mathcal{B}(\mathcal{M},\mathcal{M}')$ for P is denoted by $Win_P(\mathcal{B}(\mathcal{M},\mathcal{M}'))$.*

We have the following adequacy theorem. The proof is standard and the reader is referred to [9].

THEOREM 7 (ADEQUACY). *Take $(\mathcal{M},a)$ and $(\mathcal{M}',a')$ to be two argumentation models. It holds that:*

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

In words, **D** has a winning strategy in the total bisimulation game $\mathcal{B}(\mathcal{M},\mathcal{M}')@(a,a')$ if and only if $\mathcal{M},a$ and $\mathcal{M}',a'$ are totally bisimilar. An example of such a game follows.

EXAMPLE 4 (A TOTAL BISIMULATION GAME). *Consider two simple legal cases concerning the innocence or guiltiness of two defendants in two different trials. In the first one, two arguments a and b claiming the defendant to be guilty attack an argument a claiming his/her innocence. In the second one, only one argument x claiming the defendant's guiltiness attacks an argument y for his/her innocence. The two argumentation models, $\mathcal{M}$ and $\mathcal{M}'$, are depicted at the top of Figure 3. 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 $((\mathcal{M},b)(\mathcal{M}',y))$ by moving in the first model to state a which is labelled* `guilty` *while y is labelled* `innocent`.

Pushing the legal analogy further, bisimulation games are an idealized version of the sort of dialogues in which lawyers compare old cases with new ones. The lawyer arguing for difference proceeds like the **S**poiler, while the lawyer claiming the equivalence of the cases, proceeds like the **D**uplicator.

6. RELATED AND FUTURE WORK

6.1 Related work

To the best of our knowledge, only two papers have dealt with the application of logic to the formalization of abstract argumentation theory. The first one is [2] 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 metatheoretical results are studied.

The second one [4] 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,

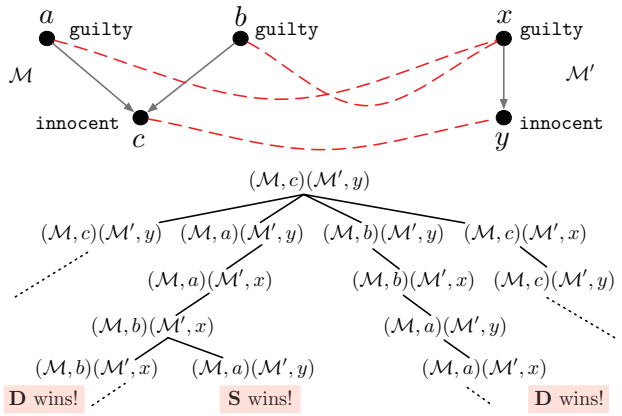

Figure 3: A total bisimulation game.

[4] proceeds from a proof-theoretic point of view, characterizing complete and grounded extensions within provability logic. Unlike in our approach, also [4] uses propositional atoms to denote arguments rather than sets thereof.

6.2 Preferred extensions in modal logic?

The paper has left aside one key notion of argumentation: preferred extensions. In [6], 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 also this notion.

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:

$$p \wedge ST_x(Compl(p)) \wedge \forall q((q \wedge ST_x(Compl(q)) \rightarrow q \sqsubseteq p) \quad (13)$$

where $ST_x(Compl(p))$ denotes the standard translation [1] of the $\mathsf{K}^{\forall}$ formula for complete extensions (Formula 4) 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 13 turns out to be invariant under total bisimulation (Definition 6).

THEOREM 8 (PREFERRED AND TOTAL BISIMULATION). *Take $\varphi(p)$ to be defined as in Formula 13 and let $\leftrightarroweq$ denote a total bisimulation relation. For any two pointed models $(\mathcal{M},a)$ and $(\mathcal{M}',a')$ it holds that:*

$$\mathcal{M},a \leftrightarroweq (\mathcal{M}',a') \implies (\mathcal{M},a) \models \varphi(p) \Longleftrightarrow \mathcal{M}',a' \models \varphi(p)$$

PROOF (SKETCH). Assume per absurdum that $\mathcal{M}',a' \models \exists q((q \wedge ST_x(Compl(q)) \wedge \neg(q \sqsubseteq p))$. By Definition 6 we obtain $\mathcal{M},a \models q \wedge ST_x(Compl(q)) \wedge \neg(q \sqsubseteq p)$ which contradicts the assumption. The other direction is similar. □

In short, Theorem 8 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 [13], Formula 13 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 8, Theorem 6 carries over to preferred extensions.

6.3 Conclusions

The paper has shown how well-known modal logics—the extensions of K with universal modality and least fixpoint operator—can be fruitfully applied to argumentation theory in an almost direct way. Future work will aim at filling the gaps in the analysis presented—eminently the formalization of preferred extensions and the study of monadic second-order games—as well as pursuing some of the research lines that this perspective opens up, such as pushing further the application of bisimulation in argumentation, and applying dynamic logic techniques to study argumentation dynamics.

Acknowledgments.

This work has been funded by the *Nederlandse Organisatie voor Wetenschappelijk Onderzoek* (NWO) under the VENI grant 639.021.816. The author wishes to thank Sanjay Modgil for the inspiring conversation that sparked this study.

7. REFERENCES

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

[2] G. Boella, J. Hulstijn, and L. van der Torre. A logic of abstract argumentation. In *Proceedings of ArgMAS 2005*, pages 29–41. Springer, 2006.

[3] M. Caminada. On the issue of reinstatement in argumentation. *Logics in Artificial Intelligence. Proceedings of JELIA 2006*, pages 111–123, 2006.

[4] M. Caminada and D. Gabbay. A logical account of formal argumentation. To appear in Studia Logica, 2009.

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

[6] 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.

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

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

[9] V. Goranko and M. Otto. Model theory of modal logic. In *Handbook of Modal Logic*, pages 249–329. Elsevier, 2007.

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

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

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

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

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

[15] 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.