When Are Two Arguments the Same? Equivalence in Abstract Argumentation

Davide Grossi and Dov Gabbay

Abstract In abstract argumentation arguments are just points in a graph of attacks: they do not hold premisses, conclusions or internal structure. So is there a meaningful way in which two arguments, belonging possibly to different attack graphs, can be said to be equivalent? The paper argues for a positive answer and, interfacing methods from modal logic, the theory of argument games and the equational approach to argumentation, puts forth and explores a formal theory of equivalence for abstract argumentation.

1 Introduction

Abstract argumentation (Dung, 1995) is the theory of structures $\langle A, \rightarrow \rangle$—called *attack graphs*—as models of the sort of conflict that occurs in argumentation, where arguments (set A) interact by attacking one another (through the binary 'attack' relation $\rightarrow$). On the one hand, this has proven to be a prolific abstraction from which to study structural properties of sets of arguments that form 'justified' or 'rational' positions in an argumentation (cf. Baroni and Giacomin (2009); Baroni et al (2011) for recent overviews). On the other hand, this perspective leaves the internal structure of arguments unspecified and arguments are nothing but points in a network of attacks. When looking at similarities between arguments from this point of view, issues such as having the same premisses and conclusions, or exhibiting the same logical structure, become immaterial.

Davide Grossi
University of Liverpool, UK, e-mail: d.grossi@liverpool.ac.uk

Dov Gabbay
King's College London, UK, e-mail: dov.gabbay@kcl.ac.uk
Bar Ilan University, Israel
University of Luxembourg, Luxembourg

However even at this level of abstraction there is a telling sense in which two arguments a and a' belonging to two (possibly different) graphs $\langle A, \rightarrow \rangle$ and $\langle A', \rightarrow' \rangle$ can be said to be 'the same', or to be equivalent, namely if they 'behave' in the same way in the two graphs. Put otherwise, a and a' can be said to be equivalent if they interact in similar ways with the other arguments in their respective graphs. This point of view suggests a way of comparing arguments which is independent of their content, and which instead stresses the role they play in an argumentation through their interaction with other arguments.

Suggestively, this 'behavioral' view of the notion of equivalence of arguments ties in well with Toulmin's view of a theory of argumentation as something that is "field-invariant":

> "What features of our arguments should we expect to be field-invariant: which features will be field-dependent? We can get some hints, if we consider the parallel between the judicial process, by which the questions raised in a law court are settled, and the rational process, by which arguments are set out and produced in support of an initial assertion. [...] One broad distinction is fairly clear. The sorts of evidence relevant in cases of different kinds will naturally be very variable. [...] On the other hand there will be, within limits, certain broad similarities between the orders of proceedings adopted in the actual trial of different cases, even when these are concerned with issues of very different kinds. [...] When we turn from the judicial to the rational process, the same broad distinction can be drawn. Certain basic *similarities of pattern and procedure* [our emphasis] can be recognized, not only among legal arguments but among justificatory arguments in general, however widely different the fields of the arguments, the sort of evidence relevant, and the weight of the evidence may be." (Toulmin, 1958, pp.15-17)

The paper aims at developing a theory of equivalence of arguments based on structural *similarities of pattern and procedure*. To this aim, the paper pushes further the application of modal logic techniques to abstract argumentation already argued for in a number of recent works (cf. Caminada and Gabbay (2009); Gabbay (2011b) and Grossi (2009, 2010, 2011)). It builds on the view of attack graphs $\langle A, \rightarrow \rangle$ as Kripke frames and presents a systematic exploration of the idea that argument equivalence can be expressed as equality of (fragments) of the modal theory of each argument. This idea naturally relates to the modal invariance notion of bisimulation (van Benthem, 1983)[1] and with the theory of argument games, that is, 'argumentation procedures' modeled as two-player zero-sum games played on attack graphs.[2] Inspired by insights from (van Benthem, 2002, 2013), we will look at a power-based notion of argument equivalence: two arguments can be said to be equivalent when the powers of the proponent and opponent in the argument games for the two arguments are, in some precise sense, the 'same'. Finally, we will see how this game-theoretic view of argumentation and argument equivalence ties in with the equational view of argumentation put forth in (Gabbay, 2011a, 2012, 2013).

Structure of the paper. In Section 2 we concisely introduce the key concepts of abstract argumentation which will be used in the paper. Section 3 provides some

[1] The relevance of bisimulation in abstract argumentation was first emphasized in (Grossi, 2009, 2010).

[2] Cf. (Modgil and Caminada, 2009) for a recent overview of argument games.

modal logic preliminaries and Section 4 applies modal equivalence to define a notion of equivalence for arguments, with respect to Dung's grounded extension. Section 5 elaborates on that definition proposing a strategic variant of it based on the powers that a proponent and an opponent have in an argument game for the grounded extension. Section 6 relates the construction of winning strategies in such argument games to the equational approach to argumentation, and brings the three strands of the paper—the modal, the game-theoretic and the equational—together. Finally, conclusions follow in Section 7.

2 Preliminaries on abstract argumentation

The present section introduces the necessary preliminaries on abstract argumentation which set the stage of our investigations.

2.1 Attack graphs

We start by the key notion of Dung (1995):

Definition 1 (Attack graph). An attack graph—or Dung framework—is a tuple $\mathscr{A} = \langle A, \rightarrow \rangle$ where:

- A is a non-empty set—the set of arguments;
- $\rightarrow \subseteq A^2$ is a binary relation—the attack relation.

The set of all attack graphs on a given set A is denoted $\mathfrak{A}(A)$. The set of all attack graphs is denoted $\mathfrak{A}$. With $a \rightarrow b$ we indicate that a attacks b, and with $X \rightarrow a$ we indicate that $\exists b \in X$ s.t. $b \rightarrow a$. Similarly, $a \rightarrow X$ indicates that $\exists b \in X$ s.t. $b \leftarrow a$. An attack graph such that, for each $a \in A$ the cardinality $|\{b \mid a \leftarrow b\}|$ of the set of the attackers of a is finite, is called *finitary*.[3] Given an argument a, we denote by $R_{\mathscr{A}}(a)$ the set of arguments attacking a: $\{b \in A \mid b \rightarrow a\}$.

These relational structures (see Figure 1 for an example) are the building blocks of abstract argumentation theory. Once A is taken to represent a set of arguments (or 'pieces of evidence' or 'information sources'), and → an 'attack' relation between arguments (so that $a \rightarrow b$ means "a attacks b"), the study of these structures provides very general insights on how competing arguments interact and structural properties of subsets of A can be taken to formalize how collections of arguments form 'justifiable' positions in an argumentation.

[3] This property is known in modal logic as *image-finiteness* of the accessibility relation of a Kripke frame (Blackburn et al, 2001, Ch. 2).

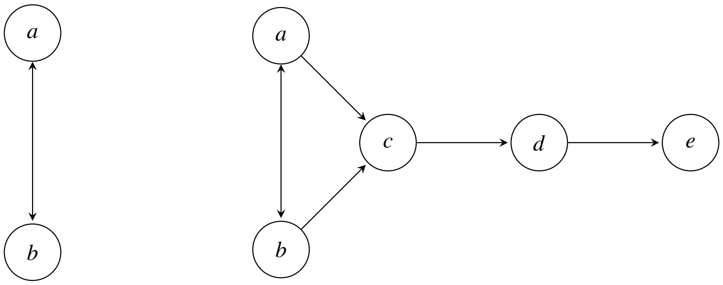

Fig. 1 Two attack graphs. The one on the left represents a full opposition between, for instance, two contradictory arguments. The one on the right represents an argumentation where two opposite arguments (*a* and *b*) both attack a same argument (*c*) which in turn defends a final argument (*e*) by attacking its attacker (*d*).

2.2 Characteristic functions of attack graphs

The formulation of all main argumentation theoretic properties makes use of two functions that can be naturally associated to each attack graph.

2.2.1 Characteristic functions

The first one is a function called in Dung (1995) characteristic function, which we will call here defense function.

Definition 2 (Defense function). Let $\mathscr{A} = \langle A, \rightarrow \rangle$ be an attack graph. The defense function $\mathtt{d}_{\mathscr{A}} : \wp(A) \longrightarrow \wp(A)$ for $\mathscr{A}$ is so defined:

$$\mathtt{d}_{\mathscr{A}}(X) = \{x \in A \mid \forall y \in A : \text{ IF } y \rightarrow x \text{ THEN } X \rightarrow y\}.$$

Given a set of arguments X, the n-fold iteration of $\mathtt{d}_{\mathscr{A}}$ is denoted $\mathtt{d}^n_{\mathscr{A}}$ for $0 \leq n < \omega$ and its infinite iteration is denoted $\mathtt{d}^{\omega}_{\mathscr{A}}$. For a given X, an infinite iteration generates an infinite sequence, or stream, $\mathtt{d}^0_{\mathscr{A}}(X), \mathtt{d}^1_{\mathscr{A}}(X), \mathtt{d}^2_{\mathscr{A}}(X), \ldots$. A stream is said to stabilize if and only if there exists $0 \leq n < \omega$ such that $\mathtt{d}^n_{\mathscr{A}}(X) = \mathtt{d}^{n+1}_{\mathscr{A}}(X)$. Such set $\mathtt{d}^n_{\mathscr{A}}(X)$ is then called the limit of the stream. When clear from the context we will drop the reference to $\mathscr{A}$ in $\mathtt{d}_{\mathscr{A}}$.

Intuitively, for a given $\mathscr{A}$, function $\mathtt{d}_{\mathscr{A}}$ encodes for each set of arguments X, which other arguments the set X is able to defend within $\mathscr{A}$.

The second function was first introduced in Pollock (1987, 1991) and further studied in Dung (1995). It is not known with a specific name in the literature. We call it here neutrality function.

Definition 3 (Neutrality function). Let $\mathscr{A} = \langle A, \rightarrow \rangle$ be an attack graph. The neutrality function $\mathtt{n}_{\mathscr{A}} : \wp(A) \longrightarrow \wp(A)$ for $\mathscr{A}$ is so defined:

$$\mathtt{n}_{\mathscr{A}}(X) = \{x \in A \mid \text{ NOT } X \to x\}$$

The definitions of n-fold iteration, stream, and stabilization are like in Definition 2.

Intuitively, given $\mathscr{A}$, function $\mathtt{n}_{\mathscr{A}}$ encodes for each set X of arguments in $\mathscr{A}$, the arguments about which X is neutral in the sense of not attacking any of those arguments.

Example 1 (Defense and neutrality in Figure 1). The functions applied to the symmetric graph on the left of Figure 1 yield the following equations:

$$\begin{array}{rcl rcl}
\mathtt{d}(\emptyset) & = & \emptyset & \mathtt{n}(\emptyset) & = & \{a,b\} \\
\mathtt{d}(\{a\}) & = & \{a\} & \mathtt{n}(\{a\}) & = & \{a\} \\
\mathtt{d}(\{b\}) & = & \{b\} & \mathtt{n}(\{b\}) & = & \{b\} \\
\mathtt{d}(\{a,b\}) & = & \{a,b\} & \mathtt{n}(\{a,b\}) & = & \emptyset
\end{array}$$

2.2.2 Properties of the defense function

We list here two properties of the defense function which will be used in the development of the paper.

The first one, monotonicity, expresses the property that larger sets of arguments are able to defend larger sets of arguments. This is enough to guarantee the existence of least and greatest fixpoints of the defense function, by the Knaster-Tarski theorem.[4]

The second one, continuity, expresses the property that in finitary graphs (i.e., graphs where arguments have at most a finite number of attackers, recall Definition 1), what is defended by a series of larger and larger sets of arguments is equivalent to the union of what each of those sets defends. As we will see later, continuity enables the possibility of studying processes of computation of argumentation-theoretic notions as iterated applications of the defense function.

Fact 1 (Monotonicity) *Let $\mathscr{A} = \langle A, \to \rangle$ be an attack graph. Function $\mathtt{n}_{\mathscr{A}}$ is monotone, i.e., for any $X, Y \subseteq A$:*

$$X \subseteq Y \Longrightarrow \mathtt{d}_{\mathscr{A}}(X) \subseteq \mathtt{d}_{\mathscr{A}}(Y).$$

Fact 2 (Continuity Dung (1995)) *Let $\mathscr{A}$ be a finitary attack graph. If $\mathscr{A}$ is finitary, then $\mathtt{d}_{\mathscr{A},X}$ is continuous for any $X \subseteq A$, i.e., for any directed set $D \in \wp(\wp(A))$: $\mathtt{d}_{\mathscr{A}}(\bigcup_{X \in D} X) = \bigcup_{X \in D} \mathtt{d}_{\mathscr{A}}(X)$.*

Proof. [RIGHT TO LEFT] Trivial. [LEFT TO RIGHT] Assume $a \in \mathtt{d}_{\mathscr{A}}(\bigcup_{X \in D} X)$. By image-finiteness there exists $X \in D$ s.t. it contains all arguments that attack some of a's attackers. Hence $a \in \bigcup_{X \in D} \mathtt{d}_{\mathscr{A}}(X)$. □

[4] The reader is referred to Davey and Priestley (1990) for a detailed presentation of this result.

X is conflict-free in $\mathscr{A}$	iff $X \subseteq \mathtt{n}_{\mathscr{A}}(X)$
X is self-defended in $\mathscr{A}$	iff $X \subseteq \mathtt{d}_{\mathscr{A}}(X)$
X is admissible in $\mathscr{A}$	iff $X \subseteq \mathtt{n}_{\mathscr{A}}(X)$ and $X \subseteq \mathtt{d}_{\mathscr{A}}(X)$
X is a complete set in $\mathscr{A}$	iff $X \subseteq \mathtt{n}_{\mathscr{A}}(X)$ and $X = \mathtt{d}_{\mathscr{A}}(X)$
X is the grounded set in $\mathscr{A}$	iff $X = \mathsf{lfp}.\mathtt{d}_{\mathscr{A}}$

Table 1 Some of the key notions of abstract argumentation theory from Dung (1995).

2.3 Solving attack graphs

By ‘solving’ an attack graph we mean selecting a subset of arguments that enjoy some characteristic structural property. The idea behind Dung’s semantics for argumentation is precisely that some structural properties of attack graphs can capture intuitive notions of justifiability of arguments or, if you wish, of standard of proof—what in argumentation are usually called *extensions*. Therefore, the study of structural properties of attack graphs delivers very general insights on how competing arguments interact and how collections of them form ‘tenable’ or ‘justifiable’ argumentative positions.

Table 1 recapitulates the basic notions of abstract argumentation which will be touching upon in the paper. They are all formulated either as fixpoints ($X = f(X)$) or post-fixpoints ($X \subseteq f(X)$) of the defense and neutrality functions, or as combinations of the two.

Intuitively, conflict-freeness demands that the set of arguments at issue is not able to attack itself—it is neutral with respect to itself. Self-defense requires that the set of arguments is able to defend itself. An admissible set is then a set of arguments which is condlict-free and is able to defend all its attackers. So, as the name suggests, admissible sets can be thought of as ‘admissible’ positions within an attack graph. By considering those admissible sets which also contain all the arguments they are able to defend—viz., the admissible sets that are fixpoints of the defense function—we obtain the notion of complete set. It formalizes the idea of a fully exploited admissible position, that is, a position which has no conflicts, and which consists exactly of all the arguments that can be successfully defended. The grounded set represents what all complete extensions have in common. In a way, it formalizes what at least must be accepted as ‘reasonable’ within the graph.

Example 2 (Extensions in Figure 1). Consider the graph on the right of Figure 1. The grounded extension is $\emptyset$. There are two complete extensions: $\{a,d\}$ and $\{b,d\}$. An example of a conflict-free set which is not admissible is $\{c,e\}$.

2.4 Computing the grounded set

We now look at a process of computation of the grounded set. This will be related later to the notion of argument equivalence to be developed, and the availability of winning strategies for the proponent in argument games.

We will focus on finitary graphs (recall Definition 1). The case of non-finitary graphs is briefly discussed in Remark 1.

Theorem 1 (Computation of grounded extensions (Dung, 1995)). *Let $\mathscr{A}$ be a finitary attack graph:*

$$\mathsf{lfp}.\mathtt{d}_{\mathscr{A}} = \bigcup_{0 \leq n < \omega} \mathtt{d}^n_{\mathscr{A}}(\emptyset) \tag{1}$$

Proof. First, we prove that $\bigcup_{0 \leq n < \omega} \mathtt{d}^n_{\mathscr{A}}(\emptyset)$ is a fixpoint by the following equations:

$$\begin{aligned} \mathtt{d}_{\mathscr{A}}\left(\bigcup_{0 \leq n < \omega} \mathtt{d}^n_{\mathscr{A}}(\emptyset)\right) &= \bigcup_{0 \leq n < \omega} \mathtt{d}_{\mathscr{A}}(\mathtt{d}^n_{\mathscr{A}}(\emptyset)) \\ &= \bigcup_{0 \leq n < \omega} \mathtt{d}^n_{\mathscr{A}}(\emptyset) \end{aligned}$$

where the first equation holds by the continuity of $\mathtt{d}_{\mathscr{A}}$, and the second since, by monotonicity, $\mathtt{d}^0_{\mathscr{A}}(\emptyset), \mathtt{d}^1_{\mathscr{A}}(\emptyset), \ldots$ is non-descending. Second, we proceed to prove that $\bigcup_{0 \leq n < \omega} \mathtt{d}^n_{\mathscr{A}}(\emptyset)$ is indeed the least fixpoint. Suppose, towards a contradiction that there exists Y s.t.: $\emptyset \subset Y = \mathtt{d}_{\mathscr{A}}(Y) \subset \bigcup_{0 \leq n < \omega} \mathtt{d}^n_{\mathscr{A}}(\emptyset)$. It follows that $\emptyset \subset Y = \mathtt{d}_{\mathscr{A}}(Y) \subset \mathtt{d}^n_{\mathscr{A}}(\emptyset)$ for some $0 \leq n < \omega$. But, by Fact 1, we have that $\mathtt{d}^n_{\mathscr{A}}(\emptyset) \subseteq \mathtt{d}^n_{\mathscr{A}}(Y)$. Contradiction. □

Remark 1 (Non-finitary graphs). For infinite graphs which are not finitary, Theorem 1 could be generalized by ordinal induction:

$$\begin{aligned} \mathtt{d}^0_{\mathscr{A}}(\emptyset) &= \emptyset \\ \mathtt{d}^{\alpha+1}_{\mathscr{A}}(\emptyset) &= \mathtt{d}_{\mathscr{A}}(\mathtt{d}^{\alpha}_{\mathscr{A}}(\emptyset)) \\ \mathtt{d}^{\lambda}_{\mathscr{A}} &= \bigcup_{\alpha < \lambda} \mathtt{d}^{\alpha}_{\mathscr{A}}(\emptyset) \quad \text{(for } \lambda \text{ arbitrary limit ordinal).} \end{aligned}$$

By the monotonicity of $\mathtt{d}_{\mathscr{A}}$ it can then be shown that there exists an ordinal α of cardinality at most the cardinal after $|A|$ such that: $\mathsf{lfp}.\mathtt{d}_{\mathscr{A}} = \mathtt{d}^{\alpha}_{\mathscr{A}}$.[5] So, in the case of non finitary attack graphs on a countable set of arguments, to obtain $\mathsf{lfp}.\mathtt{d}_{\mathscr{A}}$ we need $\bigcup_{0 \leq \alpha < \omega_1} \mathtt{d}^{\alpha}_{\mathscr{A}}(\emptyset)$.

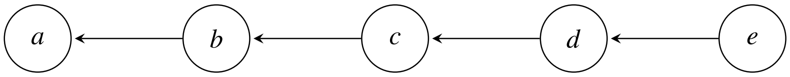

Fig. 2 A linear well-founded attack graph. The greatest and smallest fixpoint of the defense function coincide here: $\{a,c,e\}$. The set of arguments not belonging to the greatest fixpoint is $\{d,b\}$. Note, in particular, that while b is defended by set $\{a,b,c,d,e\}$ (namely by d), it is not defended by the set of arguments that is defended by $\{a,b,c,d,e\}$. So it does not belong to the greatest fixpoint of the defense function.

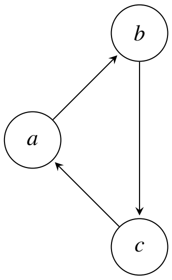

Fig. 3 A 3-cycle attack graph. Here the smallest fixpoint of the defense function is $\emptyset$ and the greatest fixpoint is $\{a,b,c\}$.

2.4.1 Smallest and greatest fixpoints of the defense function

We have seen that the smallest fixpoint of the defense function $\mathtt{d}_{\mathscr{A}}$ defines the so-called grounded extension of an attack graph. What about the largest: $\mathtt{gfp.d}_{\mathscr{A}}$? We will confine our discussion to finitary graphs.

The arguments that belong to $\mathtt{gfp.d}_{\mathscr{A}}$ are those which can always be defended by some other argument that can also in turn be defended. The dual of Theorem 1 offers a good perspective from which to appreciate the notion:

$$\mathtt{gfp.d}_{\mathscr{A}} = \bigcap_{0 \leq n < \omega} \mathtt{d}^{n}_{\mathscr{A}}(A)$$

i.e., the set consisting of arguments that are defended by the set of all arguments, and by the set that is defended by the set of all arguments and so on: $\mathtt{d}_{\mathscr{A}}(A) \cap \mathtt{d}_{\mathscr{A}}(\mathtt{d}_{\mathscr{A}}(A)) \cap \ldots$ (see Figures 2 and 3 for examples).

3 Attack graphs and modal logic

The section recapitulates and slightly extends (in particular w.r.t. frame languages) the modal logic approach to abstract argumentation put forth in Grossi (2009, 2010).

[5] A proof of this statement in the general setting of complete partial orders can be found in (Venema, 2008, Ch. 3).

3.1 Attack graphs and Kripke models

Once an attack graph is viewed as a Kripke frame, the addition of a function assigning names to sets of arguments—a labeling or valuation function—yields a Kripke model (or a state transition system).

Definition 4 (Attack models). Let $\mathbf{P}$ be a set of atoms. An attack model is a tuple $\mathscr{M} = \langle \mathscr{A}, \mathscr{V} \rangle$ where $\mathscr{A} = \langle A, \rightarrow \rangle$ is an attack graph and $\mathscr{V} : \mathbf{P} \longrightarrow \wp(A)$ is a valuation function. A pointed attack model is a pair $\langle \mathscr{M}, a \rangle$ with $a \in A$. The set of attack models is $\mathfrak{M}$.

Attack models are nothing but attack graphs together with a way of 'naming' sets of arguments or, to put it otherwise, of 'labeling' arguments.[6] So, the fact that an argument a belongs to the set $\mathscr{V}(p)$ in a given model $\mathscr{M}$ reads in logical notation as $(\mathscr{A}, \mathscr{V}), a \models p$. By using the language of propositional logic we can then form 'complex' labels φ for sets of arguments stating, for instance, that "a belongs to both the sets called p and q": $(\mathscr{A}, \mathscr{V}), a \models p \wedge q$.

In order to formalize argumentation-theoretic statements more than just propositional expressivity is needed. Let us mention a couple of examples: "*there exists* an argument in a set named φ attacking argument a" or "*for all* attackers of argument a there exist some attackers in a set named φ". These are statements involving a bounded quantification and they can be naturally formalized by a modal operator $\Diamond$ whose reading is: "there exists an attacking argument such that ...". This takes us to modal languages.

3.2 The 'being attacked' modality

Interpret now the basic modal language on argumentation models as follows:

$$\mathscr{M}, a \models \Diamond\varphi \Longleftrightarrow \exists b \in A : a \leftarrow b \text{ AND } \mathscr{M}, b \models \varphi$$

An argument a belongs to the set called $\Diamond\varphi$ iff some argument b is accessible via the inverse of the attack relation and b belongs to φ or, more simply, iff a is attacked by some argument in φ.

This is enough expressivity to express the defense and neutrality functions in modal logic K. The two functions $\mathtt{d}_{\mathscr{A}}$ and $\mathtt{n}_{\mathscr{A}}$ correspond to the functions denoted in $\mathscr{L}$ by the modal expressions $\Box\Diamond$ and, respectively, $\neg\Diamond$ on a given graph $\mathscr{A}$.

Lemma 1 (Defense and neutrality in modal logic). *Let $\mathscr{A}$ be an attack graph and $\mathscr{V}$ a valuation function.*

$$\langle \mathscr{A}, \mathscr{V} \rangle, a \models \Box\Diamond\varphi \Longleftrightarrow a \in \mathtt{d}_{\mathscr{A}}([\![\varphi]\!]_{\langle \mathscr{A}, \mathscr{V} \rangle})$$
$$\langle \mathscr{A}, \mathscr{V} \rangle, a \models \neg\Diamond\varphi \Longleftrightarrow a \in \mathtt{n}_{\mathscr{A}}([\![\varphi]\!]_{\langle \mathscr{A}, \mathscr{V} \rangle})$$

[6] It might be worth noticing that this is a generalization of the sort of labeling functions studied in argumentation theory (cf. Caminada (2006); Baroni and Giacomin (2009)).

Proof (Sketch of proof). For $\Box\Diamond$ we have these equivalences:

$$\begin{aligned}\llbracket \Box\Diamond\varphi \rrbracket_{\langle \mathscr{A},\mathscr{V}\rangle} &= \{a \mid \forall b : \text{ IF } a \leftarrow b \text{ THEN } b \leftarrow \llbracket \varphi \rrbracket_{\langle \mathscr{A},\mathscr{V}\rangle}\} \\ &= \mathtt{d}_{\mathscr{A}}(\llbracket \varphi \rrbracket_{\langle \mathscr{A},\mathscr{V}\rangle}).\end{aligned}$$

The first equation holds by construction, the second and third are application of the the semantics of $\Box\Diamond$ and Definition 2. The reasoning for $\neg\Diamond\varphi$ is analogous.[7] □

In general, emphasizing the modal nature of $\mathtt{d}_{\mathscr{A}}$ and $\mathtt{n}_{\mathscr{A}}$ has the advantage of allowing us to use available modal principles in reasoning about argumentation-theoretic notions. All the theorems of logic K concerning $\Box\Diamond$- and $\neg\Diamond$-formulae can legitimately be seen as theorems of abstract argumentation. Here we list a few very simple theorems of K which carry interesting readings in terms of abstract argumentation theory.

Fact 3 *The following are theorems of* K*:*

$$\Box\Diamond\bot \leftrightarrow \Box\bot \tag{2}$$

$$\Box\Diamond\varphi \leftrightarrow \neg\Diamond\neg\Diamond\varphi \tag{3}$$

$$\Box\Diamond\Box\Diamond\bot \leftrightarrow \Box\Diamond\bot \vee \Box\Diamond\Box\Diamond\bot \tag{4}$$

Formula 2 uses the trivial modal fact that $\Diamond\bot \leftrightarrow \bot$ to express that the set of arguments defended by the empty set corresponds to the set of arguments that have no attackers (dead ends). This equivalence will be constantly used in the remainder of the paper. Formula (3) is the modal counterpart of the equivalence of the defense function and the 2-fold iteration of the neutrality function, i.e., for any X and graph $\mathscr{A}$: $\mathtt{d}_{\mathscr{A}}(X) = \mathtt{n}_{\mathscr{A}}(\mathtt{n}_{\mathscr{A}}(X))$. Formula (4) states that, for any $\mathscr{A}$, the finite union of subsequent iterations of $\mathtt{d}_{\mathscr{A}}$ over $\emptyset$ is equivalent to the longest iteration.

In the remainder of the paper, in order to concisely express the n^{th} iteration of $\Box\Diamond$ (resp., $\neg\Diamond$) we will write $(\Box\Diamond)^n$ (resp., $(\neg\Diamond)^n$).

Remark 2 (Frame language). When interested in the application of the characteristic functions solely to the set of all arguments, or to the empty set of arguments, all is needed to express $\mathtt{d}$ and $\mathtt{n}$ is a limited fragment of the language $\mathscr{L}$ introduced above. The fragment is defined by the following BNF:

$$\varphi ::= \bot \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Diamond\varphi$$

This is a so-called *frame language*[8], which does not use propositional atoms. In fact, this language does not need models to be interpreted, but simply attack graphs (Definition 1). It therefore expresses properties of pointed attack graphs: $\langle \mathscr{A}, a\rangle$.

[7] More generally, the claim is a direct consequence of the existence of a homomorphism from the term algebra $\mathsf{Term} = \langle \mathscr{L}, \wedge, \neg, \bot, \Diamond\rangle$ of language $\mathscr{L}$ (without universal modality) to the complex algebra $\mathsf{Set}_{\mathscr{A}} = \langle 2^A, \cap, -, \emptyset, f\rangle$ where $f : \wp(A) \longrightarrow \wp(A)$ such that $f(A) = \{a \in A \mid \exists b \in A : a \leftarrow b\}$ (Blackburn et al, 2001, Ch. 5).

[8] See (Blackburn et al, 2001, Ch. 3.1).

This will be the language we will be working with when defining a notion of argument equivalence with respect to the grounded set.

3.2.1 The grounded set in modal logic

As a consequence of Theorem 1 and Lemma 1—showing that $\mathtt{d}$ can be represented as $\Box\Diamond$—the grounded extension can, in any finitary graph $\mathscr{A}$, be expressed by the following infinite but countable disjunction (cf. Equation (1)):

$$\bigvee_{0 \leq n < \omega} (\Box\Diamond)^n \bot \tag{5}$$

Clearly, in a finite $\mathscr{A}$ we will have a finite integer n where the stream $\mathtt{d}^{\omega}_{\mathscr{A}}(\emptyset)$ reaches its limit, and we could then express the grounded extension by a finite disjunction $\bigvee_{0 \leq i \leq n} (\Box\Diamond)^i \bot$ or simply as $(\Box\Diamond)^i \bot$.

Similarly, it is worth observing that the greatest fixpoint of $\mathtt{d}_{\mathscr{A}}$ for a given $\mathscr{A}$ is expressed by the following infinite conjunction:

$$\bigwedge_{0 \leq n < \omega} \neg(\Diamond\Box)^n \bot \tag{6}$$

i.e., it is neither the case that the current argument is attacked by a dead end, nor that it is attacked by an argument whose attackers are attacked by a dead end, and so on.

Remark 3 (Being attacked by the grounded set). Notice that arguments not belonging to the greatest fixpoint of $\mathtt{d}$, i.e., satisfying $\neg \bigwedge_{0 \leq n < \omega} \neg(\Diamond\Box)^n \Diamond\bot$, are arguments attacked by the grounded set, i.e., arguments satisfying $\bigvee_{0 \leq n < \omega} \Diamond(\Box\Diamond)^n \bot$.

Remark 4 (Infinite attack graphs and the mu-calculus). In the general case, in order to express the grounded extension modally it is necessary to resort to the expressivity of the mu-calculus, where the grounded extension can be expressed by the following formula:

$$\mu p.\Box\Diamond p \tag{7}$$

denoting precisely the smallest fixpoint of function $\Box\Diamond$, i.e., in a given $\mathscr{A}$, the modal rendering of $\mathtt{d}_{\mathscr{A}}$ (Lemma 1). Similarly, $\nu p.\Box\Diamond p$ denotes the largest fixpoint. We refer the reader to (Grossi, 2010; Gratie et al, 2012) for more information on the application of the modal mu-calculus to abstract argumentation.

3.2.2 Other argumentation-theoretic notions in modal logic

We have shown how to express the grounded extension by a formula of the basic frame language. It must be clear that, from a modal point of view, the grounded extension is therefore a property of a pointed frame $\langle \mathscr{A}, a \rangle$, that is, the property of an argument in a graph.

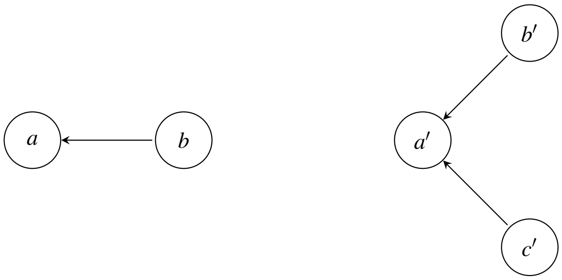

Fig. 4 Arguments a and a' have the same status: $\mathbf{T}(a) = \emptyset = \mathbf{T}(a')$.

How are the other notions of Table 1 to be formalized? In (Grossi, 2010) it has been shown that logic K with the universal modality $\langle \mathsf{U} \rangle$ suffices to express conflict-freeness, self-defense, admissibility and complete extensions. But in this case, the full modal language (with at least one atom p) is required:

$$\begin{aligned}
\mathcal{V}(p) \text{ is conflict-free} &\iff \langle \mathcal{A}, \mathcal{V} \rangle, a \models [\mathsf{U}](p \to \neg \Diamond p) \\
\mathcal{V}(p) \text{ is self-defended} &\iff \langle \mathcal{A}, \mathcal{V} \rangle, a \models [\mathsf{U}](p \to \Box \Diamond p) \\
\mathcal{V}(p) \text{ is admissible} &\iff \langle \mathcal{A}, \mathcal{V} \rangle, a \models [\mathsf{U}](p \to \neg \Diamond p) \wedge [\mathsf{U}](p \to \Box \Diamond p) \\
\mathcal{V}(p) \text{ is a complete set} &\iff \langle \mathcal{A}, \mathcal{V} \rangle, a \models [\mathsf{U}](p \to \neg \Diamond p) \wedge [\mathsf{U}](p \leftrightarrow \Box \Diamond p)
\end{aligned}$$

These notions are therefore properties of pointed models $\langle \mathcal{M}, a \rangle$, that is, properties of arguments in a graph where a set of arguments has been labeled. In the remainder of the paper we will be concerned only with frame properties and will therefore be working with the frame language.

4 A modal notion of argument equivalence

The section develops a modal notion of argument equivalence characterizing the status of an argument in terms of a special family of modal formulae it satisfies.

4.1 When are two arguments equivalent w.r.t. the grounded set?

Let us start by recalling a few observations from Section 3. For any graph $\mathcal{A}$, the set of arguments is partitioned in the set of arguments belonging to $\mathsf{lfp.d}_{\mathcal{A}}$ (the grounded set), those not belonging to $\mathsf{gfp.d}_{\mathcal{A}}$ (i.e., the arguments attacked by the grounded set, recall Remark 3), and the arguments belonging to $\mathsf{gfp.d}_{\mathcal{A}} - \mathsf{lfp.d}_{\mathcal{A}}$ (i.e., the arguments neither belonging to the grounded set nor being attacked by it). Figures 2 and 3 offer good examples for the identification of this tripartition.

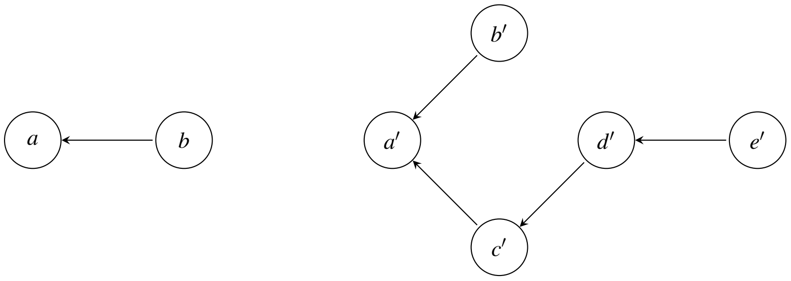

Fig. 5 Arguments b and c' have different statuses: $\mathbf{T}(b) = \{(\Box\Diamond)^n\bot \mid 1 \leq n < \omega\} \cup \{(\Box\Diamond)^n\top \mid 0 \leq n < \omega\}$, while $\mathbf{T}(c') = \{(\Box\Diamond)^n\bot \mid 2 \leq n < \omega\} \cup \{(\Box\Diamond)^n\top \mid 0 \leq n < \omega\}$. Both belong to the grounded sets of the respective graphs.

So, from the point of view of the grounded set, what matters in a graph $\mathscr{A}$ is the status of an argument with respect to the three above sets, and hence with respect to membership to $\mathsf{lfp.d}_{\mathscr{A}}$ and $\mathsf{gfp.d}_{\mathscr{A}}$. A natural refinement of this idea in finitary graphs is to understand the status of an argument not only in terms of its membership to $\mathsf{lfp.d}_{\mathscr{A}}$ and $\mathsf{gfp.d}_{\mathscr{A}}$, but also in terms of 'when' it enters those sets, in the sense of which are the stages in the fixpoint computation to which the argument belongs[9], i.e., at which n the argument comes to belong to $\mathsf{d}^n_{\mathscr{A}}(\emptyset)$ and at which it ceases to belong to $\mathsf{d}^n_{\mathscr{A}}(A)$. This suggests the following definition of status of an argument:

Definition 5 (Status). Let $\mathscr{A} = \langle A, \rightarrow\rangle$ be an attach graph. The status of $a \in A$ is defined as, for $1 \leq n < \omega$:

$$\mathbf{T}(a) = \{(\Box\Diamond)^n\bot \mid \mathscr{A}, a \models (\Box\Diamond)^n\bot\} \cup \{(\Box\Diamond)^n\top \mid \mathscr{A}, a \models (\Box\Diamond)^n\top\} \tag{8}$$

Recall the modal principle: $(\Box\Diamond)^n\top \leftrightarrow \neg\Diamond(\Box\Diamond)^n\bot$. So, the status of an argument is the subset of its modal theory in the frame language which consists of formulae corresponding to iterations of the defense function over $\emptyset$ (i.e., $\bot$) and over the set of all arguments (i.e., $\top$).

To familiarize ourselves with the notion of argument status, let us mention this simple fact following from Theorem 1:

Fact 4 *Let $\mathscr{A}$ be a finitary graph:*

$$\begin{aligned}
a \in \mathsf{lfp.d}_{\mathscr{A}} &\iff \mathbf{T}(a) = \{(\Box\Diamond)^m\bot \mid \exists n : n \leq m < \omega\} \cup \{(\Box\Diamond)^n\top \mid 1 \leq n < \omega\} \\
a \in -\mathsf{gfp.d}_{\mathscr{A}} &\iff \mathbf{T}(a) = \{(\Box\Diamond)^m\top \mid \exists n : m \leq n < \omega\} \\
a \in \mathsf{gfp.d}_{\mathscr{A}} - \mathsf{lfp.d}_{\mathscr{A}} &\iff \mathbf{T}(a) = \{(\Box\Diamond)^n\top \mid 1 \leq n < \omega\}
\end{aligned}$$

We can then say that two arguments are equivalent w.r.t. the grounded set (notation, $\mathscr{A}, a \equiv_{\mathsf{d}} \mathscr{A}', a'$) if and only if they have the same status:

[9] It is worth stressing that this is a refinement of the common understanding of 'status of an argument' in the literature on argumentation theory.

$$\mathscr{A}, a \equiv_{\mathtt{d}} \mathscr{A}', a' \Longleftrightarrow \mathbf{T}(a) = \mathbf{T}(a') \tag{9}$$

Intuitively, two arguments are equivalent if and only if they belong to exactly the same stages of iteration of the defense function applied to the empty set, and to the same stages of iteration of the defense function applied to the set of all arguments. Figures 4 and 5 give an illustration of the definitions of status and status equivalence.

4.2 Status equivalence and frame bisimulation

We recall the standard definition of the notion of frame bisimulation:[10]

Definition 6 (Frame bisimulation (van Benthem, 1983)). Let $\mathscr{A} = \langle A, \rightarrow \rangle$ and $\mathscr{A}' = \langle A', \rightarrow' \rangle$ be two attack graphs. A (frame-)bisimulation between $\mathscr{A}$ and $\mathscr{A}'$ is a non-empty relation $Z \subseteq A \times A'$ such that:

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

When a frame bisimulation exists linking $a \in A$ and $a' \in A'$ we write $\mathscr{A}, a \,\underline{\leftrightarrow}\, \mathscr{A}', a'$.

Intuitively, a bisimulation is a process-like view of equivalence between attack graphs that links the walks along the attack relation—one might say the dialogues (cf. Section 5)—that one can do on one graphs to corresponding walks that one can do on the other.

By applying standard results from modal logic we can show that frame bisimulation implies status equivalence: two bisimilar arguments are also equivalent with respect to their status.

Fact 5 **($\underline{\leftrightarrow} \subseteq \equiv_{\mathtt{d}}$)** *Let $\langle \mathscr{A}, a \rangle$ and $\langle \mathscr{A}', a' \rangle$ be two pointed attack graphs:*

$$\mathscr{A}, a \,\underline{\leftrightarrow}\, \mathscr{A}', a' \Longrightarrow \mathscr{A}, a \equiv_{\mathtt{d}} \mathscr{A}', a'$$

Proof. The claim is a direct consequence of Formula (9) and the fact that the basic modal language is bisimulation invariant (cf. Blackburn et al (2001)).

5 Status equivalence and argument games

The picture of argumentation we have given so far is of a static kind, but argumentation calls intuitively for a process of interaction between arguers. In fact, although notions like the grounded extension formalize different static views of what makes a set of arguments a 'justifiable' or good position in an argumentation, these views can be made dynamic through two-player zero-sum games. Many researchers in the

[10] See (Blackburn et al, 2001, Ch. 2).

Length of $\mathbf{a}$:	$\mathscr{P}$ wins if :	$\mathscr{O}$ wins if :
$\ell(\mathbf{a}) < \omega$	$\mathtt{t}(\mathbf{a}) = \mathscr{O}$	$\mathtt{t}(\mathbf{a}) = \mathscr{P}$
$\ell(\mathbf{a}) = \omega$	*never*	*always*

Table 2 Winning conditions for the game for grounded given a terminal dialogue **a**.

last two decades have focused on 'dialogue games' for argumentation, i.e., games able to adequately establish whether a given argument belongs or not to a given extension.[11]

The sort of results that drive this literature are called *adequacy* theorems and have, roughly, the following form: argument a has property S (e.g., belongs to the grounded extension) if and only if the proponent has a winning strategy in the dialogue game for property S (e.g., the dialogue game for the grounded extension) starting with argument a.

In this section we will see how the notion of bisimulation between arguments ties in with the theory of argument games.

5.1 Argument games

The section recapitulates key definitions and results pertaining to an adequate game for the grounded extension.

5.1.1 Game for the grounded extension

Let us fix some further auxiliary notation before starting. Let $\mathbf{a} \in A^{<\omega} \cup A^{\omega}$ be a finite or infinite sequence of arguments in A, which we will call a *dialogue*. To denote the n^{th} element, for $1 \leq n < \omega$, of a dialogue $\mathbf{a}$ we write $\mathbf{a}_n$, and to denote the dialogue consisting of the first n elements of $\mathbf{a}$ we write $\mathbf{a}|_n$. The last argument of a finite dialogue $\mathbf{a}$ is denoted $h(\mathbf{a})$. Finally, the length $\ell(\mathbf{a})$ of $\mathbf{a}$ is $n-1$ if $\mathbf{a}|_n = \mathbf{a}$, and ω otherwise. We start with the formal definition:

Definition 7 (Argument game for grounded (Dung, 1994)). The game for the grounded extension is a function $\mathscr{D}$ which for each attack graph $\mathscr{A}$ yields a structure $\mathscr{D}(\mathscr{A}) = \langle N, A, \mathtt{t}, \mathtt{m}, \mathtt{p} \rangle$ where:

- $N = \{\mathscr{P}, \mathscr{O}\}$—the set of players consists of proponent $\mathscr{P}$ and opponent $\mathscr{O}$.
- A is the set of arguments in $\mathscr{A}$.

[11] The contributions that started this line of research is Dung (1994). Cf. Modgil and Caminada (2009) for a recent overview.

- $\mathtt{t} : A^{<\omega} \longrightarrow N$ is the *turn function*. It is a (partial[12]) function assigning one player to each finite dialogue in such a way that, for any $0 \leq m < \omega$ and $\mathbf{a} \in A^{<\omega}$, if $\ell(\mathbf{a}) = 2m$ then $\mathtt{t}(\mathbf{a}) = \mathscr{O}$, and if $\ell(\mathbf{a}) = 2m+1$ then $\mathtt{t}(\mathbf{a}) = \mathscr{P}$. I.e., even positions are assigned to $\mathscr{O}$ and odd positions to $\mathscr{P}$.
- $\mathtt{m} : A^{<\omega} \longrightarrow \wp(A)$ is a (partial) function from dialogues to sets of arguments defined as: $\mathtt{m}(\mathbf{a}) = R_{\mathscr{A}}(h(\mathbf{a}))$. I.e., the available moves at $\mathbf{a}$ are the arguments attacking the last argument of $\mathbf{a}$. The set of all dialogues compatible with $\mathtt{m}$—the legal dialogues of the game—is denoted D. Dialogues $\mathbf{a}$ for which $\mathtt{m}(\mathbf{a}) = \emptyset$ or such that $\ell(\mathbf{a}) = \omega$ are called terminal, and the set of all terminal dialogues of the game is denoted T.
- $\mathtt{p} : T \longrightarrow N$ is the payoff function given in Table 2, which associates a player—the winner—to each terminal dialogue.

The game is played starting from a given argument a. When a is explicitly given we talk about an instantiated game (notation, $\mathscr{D}(\mathscr{A})@a$).

The two players play the game by alternating each other ($\mathscr{O}$ starts) and navigating the attack graph along the 'being attacked' relation. The winning conditions state that $\mathscr{P}$ wins whenever she manages to state an argument to which $\mathscr{O}$ cannot reply, i.e., an argument with no attackers. Notice the asymmetry in the winning conditions of the payoff function for $\mathscr{P}$ and $\mathscr{O}$.

5.1.2 Adequacy

The different ways in which proponent and opponent can play an argument game are called strategies:

Definition 8 (Strategies). Let $\mathscr{D}(\mathscr{A}) = \langle N, A, \mathtt{t}, \mathtt{m}, \mathtt{p} \rangle$, $a \in A$ and $i \in N$. A strategy for i in the instantiated game $\mathscr{D}@a$ is a function: $\sigma_i : \{\mathbf{a} \in D - T \mid \mathbf{a}_0 = a \text{ AND } \mathtt{t}(\mathbf{a}) = i\} \longrightarrow A$ s.t. $\sigma_i(\mathbf{a}) \in \mathtt{m}(\mathbf{a})$. The set of terminal dialogues compatible with σ_i is defined as follows: $T_{\sigma_i} = \{\mathbf{a} \in T \mid \mathbf{a}_0 = a \text{ AND } \forall n \leq \ell(\mathbf{a}) \text{ IF } \mathtt{t}(\mathbf{a}|_n) = i \text{ THEN } \mathbf{a}_{n+1} = \sigma_i(\mathbf{a}|_n)\}$.

A strategy tells i which argument to choose, among the available ones, at each non-terminal dialogue $\mathbf{a}$ in $\mathscr{D}@a$. So, in the game for grounded, a strategy $\sigma_{\mathscr{P}}$ will encode the proponent's choices in dialogues of odd length, while $\sigma_{\mathscr{O}}$ will encode the opponent's choices in dialogues of even length. Observe that, in a game for grounded, a strategy $\sigma_{\mathscr{P}}$ and a strategy $\sigma_{\mathscr{O}}$—i.e., a strategy profile in the game-theory terminology—together determine one terminal dialogue or, in other words, $T_{\sigma_{\mathscr{P}}} \cap T_{\sigma_{\mathscr{O}}}$ is a singleton.

What matters of a strategy is whether it will guarantee the player that plays according to it to win the game. This brings us to the notion of winning strategy:

[12] The function is partial because only sequences compatible with the move function $\mathtt{m}$ need to be considered.

Definition 9 (Winning strategies and arguments). Let $\mathscr{D}(\mathscr{A}) = \langle N, S, \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$, $a \in A$ and $i \in N$. A strategy σ is *winning for i* in $\mathscr{D}(\mathscr{A})@a$ if and only if for all $\mathbf{a} \in T_\sigma$ it is the case that $\mathtt{p}(\mathbf{a}) = i$. An argument a is *winning for i* iff there exists a winning strategy for i in $\mathscr{D}(\mathscr{A})@a$. The set of winning positions of $\mathscr{D}$ for i is denoted $Win_i(\mathscr{D}(\mathscr{A}))$. An argument a is *winning for i in k rounds* ($k \geq 0$) iff there exists a winning strategy σ_i in $\mathscr{D}@a$ such that for all $\mathbf{a} \in T_{\sigma_i}$, $\ell(\mathbf{a}) + 1 \leq k$, that is, i can always win in at most k rounds using σ_i. The set of winning positions in k rounds is denoted $Win^k_i(\mathscr{D})$.

Dialogue games are two-player zero-sum games with perfect information. It follows that these games are determined (Zermelo's theorem), in the sense that either $\mathscr{P}$ or $\mathscr{O}$ possesses a winning strategy, and hence that each argument in an attack graph is either a winning position for $\mathscr{P}$ or a winning position for $\mathscr{O}$. See Figure 6 for an illustration.

Now all ingredients are in place to study the property we are interested in, viz. the adequacy of the game of Definition 7 with respect to the grounded extension. We first prove a slightly stronger result: an argument a belongs to the k^{th} iteration of the defense function on the empty set of arguments, if and only if $\mathscr{P}$ has a winning strategy in the game initiated at a, which she can carry out in at most $2(k-1)$ rounds.

Lemma 2 (Strong adequacy of the game for grounded (Dung, 1994)). *Let $\mathscr{D}(\mathscr{A})$ be the dialogue game for grounded on graph $\mathscr{A}$ and $a \in A$, for $1 \leq k < \omega$:*

$$a \in \mathtt{d}^k_{\mathscr{A}}(\emptyset) \iff a \in Win^{2(k-1)}_{\mathscr{P}}(\mathscr{D}(\mathscr{A})).$$

Proof. We proceed by induction:
[B] The following equivalences prove the induction base:

$$\begin{aligned} a \in \mathtt{d}_{\mathscr{A}}(\emptyset) &\iff \nexists b : b \rightarrow a && \text{[Definition 2]} \\ &\iff a \in Win^0_{\mathscr{P}}(\mathscr{D}(\mathscr{A})) && \text{[Definition 9]} \end{aligned}$$

[S] If $a \in \mathtt{d}^n_{\mathscr{A}}(\emptyset) \iff a \in Win^{2(n-1)}_{\mathscr{P}}(\mathscr{D}(\mathscr{A}))$ (IH) then we claim: $a \in \mathtt{d}^{n+1}_{\mathscr{A}}(\emptyset) \iff a \in Win^{2n}_{\mathscr{P}}(\mathscr{D}(\mathscr{A}))$. [LEFT TO RIGHT] Assume $a \in \mathtt{d}^{n+1}_{\mathscr{A}}(\emptyset)$. This means that $\forall b : b \rightarrow a, \exists c : c \rightarrow b$ and such that $c \in \mathtt{d}^n_{\mathscr{A}}(\emptyset)$ which, by IH, is equivalent to $c \in Win^{2(n-1)}_{\mathscr{P}}(\mathscr{D}(\mathscr{A}))$. So, by Definition 7, for any $\mathscr{O}$'s move b at position a, $\mathscr{P}$ has a counter-argument c from which she has a winning strategy forcing a win in at most $2(n-1)$ rounds. Hence, by Definition 9, $\mathscr{P}$ can win the game at a in $2(n-1)+2$ rounds, i.e., $a \in Win^{2n}_{\mathscr{P}}(\mathscr{D}(\mathscr{A}))$. [RIGHT TO LEFT] Assume $a \in Win^{2n}_{\mathscr{P}}(\mathscr{D}(\mathscr{A}))$. This means that, for any $\mathscr{O}$'s move b at a, $\mathscr{P}$ has a counter-argument c from which she has a winning strategy forcing a win in at most $2n-2$ rounds. By IH, this is equivalent with $c \in \mathtt{d}^n_{\mathscr{A}}(\emptyset)$ and by Definition 2 we conclude that $a \in \mathtt{d}^{n+1}_{\mathscr{A}}(\emptyset)$. This completes the proof.

As a consequence, an argument belongs to the grounded extension of an argumentation framework if and only if the proponent has a winning strategy for the dialogue game for grounded (in that argumentation framework) instantiated at that argument.

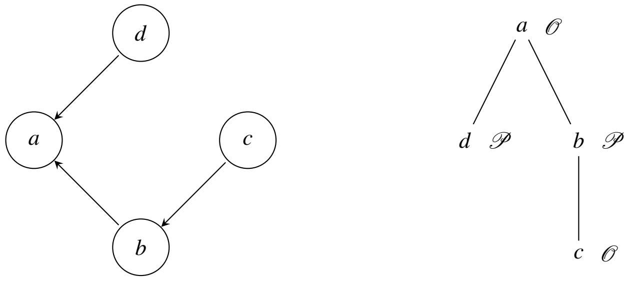

Fig. 6 An attack graph (left) and its dialogue game for grounded (right). Positions are labeled by the player whose turn it is to play. $\mathscr{P}$ wins the terminal dialogue *abc* but loses the terminal dialogue *ad*. $\mathscr{O}$ has a winning strategy that makes him win in one move.

Theorem 2 (Adequacy of the game for grounded). *Let* $\mathscr{D}(\mathscr{A}) = \langle N, S, \mathtt{t}, \mathtt{m}, \mathtt{p} \rangle$ *be the dialogue game for grounded on a finitary graph* $\mathscr{A}$ *and* $a \in A$*:*

$$a \in \mathsf{lfp}.\mathtt{d}_{\mathscr{A}} \iff a \in Win_{\mathscr{P}}(\mathscr{D}(\mathscr{A})).$$

Proof. The claim is proven by the following series of equivalences:

$$\begin{aligned} a \in Win_{\mathscr{P}}(\mathscr{D}(\mathscr{A})) &\iff a \in \bigcup_{1 \leq k < \omega} Win^{2(k-1)}_{\mathscr{P}}(\mathscr{D}(\mathscr{A})) \\ &\iff a \in \bigcup_{1 \leq k < \omega} \mathtt{d}^{k}_{\mathscr{A}}(\emptyset) \\ &\iff a \in \mathsf{lfp}.\mathtt{d}_{\mathscr{A}} \end{aligned}$$

The first equivalence holds by the winning conditions of Definition 7 and Definition 9: $\mathscr{P}$ wins if and only if she can force the game to reach an unattacked argument in an even number of steps. The second equivalence holds by Lemma 2 and the third one by Theorem 1.

Theorems like Theorem 2 play a significant role in the development of a formal theory of argumentation. Firstly, they guarantee that the argument game at issue is a sound (if the proponent has a winning strategy then the the argument is grounded) and complete (if the argument is grounded, then the proponent has a winning strategy) proof procedure with respect to the corresponding semantics. Secondly, literature in argumentation (e.g., Atkinson and Bench-Capon (2007)) has pointed out—convincingly in our view—that Dung's extensions can be soundly viewed as abstract models of standards of proof in debates, and that argument games are a viable abstraction of procedural rules, or protocols, for debates. (cf. Prakken (2009)). Viewed in this light, adequacy is then the property of debate protocols successfully *implementing* a given standard of proof, like the grounded extension.[13]

[13] We use the word "implement" here in the technical sense in which it is typically used in game theory (Osborne and Rubinstein, 1994, Ch. 10) or social software (Parikh, 2002).

5.2 Strategic equivalence of arguments & status equivalence

When can two arguments in two attack graphs be considered equivalent from the point of view of the above game? Intuitively, we might say that the two arguments are equivalent if the proponent (respectively, the opponent) has a winning strategy that allows her (respectively, him) to win the game in at most the same number of rounds. More precisely:

Definition 10 (Strategic equivalence of arguments). Two pointed attack graphs $\langle \mathscr{A}, a\rangle$ and $\langle \mathscr{A}', a'\rangle$ are *strategically equivalent* if and only if the two following conditions are met:

(i) For $0 \leq n < \omega$, if $\mathscr{P}$ can always win $\mathscr{D}(\mathscr{A})@a$ in at most $2n$ rounds, then she can always win $\mathscr{D}(\mathscr{A}')@a'$ in at most the same number of rounds, and vice versa;

(ii) For $0 \leq n < \omega$, if $\mathscr{O}$ can always win $\mathscr{D}(\mathscr{A})@a$ in at most $2n+1$ rounds, then he can always win $\mathscr{D}(\mathscr{A}')@a'$ in at most the same number of rounds, and vice versa.

In other words, two arguments are strategically equivalent whenever they support the same 'powers' for the proponent and the opponent, that is, whenever they support winning strategies (for the proponent or the opponent) that can force a win in the game for grounded in at most the same number of rounds.

Example 3. Let us get back to Figure 5. Consider arguments a and a'. These are strategically equivalent: $\mathscr{O}$ has a winning strategy for the arguments, on both graphs, guaranteeing him a win in at most 1 round. Consider now arguments b and c'. $\mathscr{P}$ has a winning strategy on both games. But while she always wins in 0 rounds from b, she always wins in 2 rounds playing from c'. So, b and c' are not strategically equivalent.

Now, capitalizing on Lemmata 1 and 2, this notion of strategic equivalence can be shown to be just a game-theoretic variant of the notion of status equivalence:

Theorem 3. *Let $\langle \mathscr{A}, a\rangle$ and $\langle \mathscr{A}', a'\rangle$ be two pointed attack graphs: $\mathscr{A}, a \equiv_{\mathtt{d}} \mathscr{A}', a'$ if and only if $\langle \mathscr{A}, a\rangle$ and $\langle \mathscr{A}', a'\rangle$ are strategically equivalent.*

Proof. Define the following set:

$$W(a) = \left\{ \begin{array}{c} \left\{ (\Box\Diamond)^n \bot \mid a \in Win_{\mathscr{P}}^{2(n-1)}(\mathscr{D}(\mathscr{A})), \text{ FOR } 1 \leq n < \omega \right\} \\ \cup \\ \left\{ (\Box\Diamond)^n \top \mid \nexists b : a \leftarrow b \text{ AND } b \in Win_{\mathscr{P}}^{2(n-1)}(\mathscr{D}(\mathscr{A})), \text{ FOR } 1 \leq n < \omega \right\} \end{array} \right.$$

First of all, observe that, for any n, if $\exists b : a \leftarrow b$ AND $b \in Win_{\mathscr{P}}^{2(n-1)}(\mathscr{D}(\mathscr{A})$ then $\mathscr{O}$ has a winning strategy in a that forces a win in $2(n-1)+1$ rounds (in symbols, $a \in Win_{\mathscr{O}}^{2n-1}(\mathscr{D}(\mathscr{A})))$, and vice versa. So, by Lemmata 1 and 2, it is not difficult to see that $\langle \mathscr{A}, a\rangle$ and $\langle \mathscr{A}', a'\rangle$ are strategically equivalent if and only if $W(a) = W(a')$ (recall that $(\Box\Diamond)^n\top \leftrightarrow \neg\Diamond(\Box\Diamond)^n\bot$). By the definition of W, Definition 5 and Lemma 1 it then follows directly that $\mathbf{T}(a) = \mathbf{T}(a')$.

We have thus shown that the modally defined notion of status equivalence for the grounded extension has a natural strategic variant based on the argument game for that extension. As a direct consequence of Fact 5 we also obtain that if two arguments are frame bisimilar, then they are strategically equivalent.

Getting back to the Toulmin's quote by which we opened the paper, Theorem 3 establishes an equivalence of arguments in terms of a procedural equivalence relating the ways proponent and opponent are able to argue with respect to the argument at issue. Two arguments in two different argumentations can be said to be equivalent whenever the powers—intended as the availability of a strategy to force a win in a fixed number of rounds—of the proponent and the opponent in the two graphs are the same. This ties in well with power-based notions of game equivalence as put forth, for instance, in (van Benthem, 2002, 2013).

6 Games and equations

In this final section we look at one more perspective on argument equivalence, based on the equational semantics of abstract argumentation (Gabbay, 2011a).

6.1 The equational approach to abstract argumentation

Let us start with a few preliminaries. The equational approach to—or equational semantics of—argumentation consists in extracting from a given finite attack graph $\mathscr{A} = \langle A, \rightarrow\rangle$ a system of equations:

$$\begin{aligned} f(a_1) &= 1 - \max(\{f(b) \mid a_1 \leftarrow b\}) \\ f(a_2) &= 1 - \max(\{f(b) \mid a_2 \leftarrow b\}) \\ \ldots &\quad \ldots \\ f(a_n) &= 1 - \max(\{f(b) \mid a_n \leftarrow b\}) \end{aligned}$$

where $a_1, \ldots, a_n$ is an enumeration of the arguments in A, and $f : A \longrightarrow [0,1]$ is a function from the sets of arguments to the real values between 0 and 1.[14] Intuitively, 0 represents a form of rejection of the argument, 1 a form of acceptance, and intermediate values a form of undecidedness.

[14] Other systems making use of different mathematical functions instead of $1 - \max(.)$ are discussed in Gabbay (2011a). See also Gabbay (2012) for an extensive exposition of the equational approach to argumentation.

As shown in Gabbay (2011a), each solution f to one such system of equations defines a set of arguments $\{a \in A \mid f(a) = 1\}$ corresponding to a complete extension (see Table 1) of the underlying attack graph. The solution f_g such that $\{a \mid f_g(a) = 1\}$ is minimal corresponds therefore to the grounded extension, i.e., to $\mathsf{lfp.d}_{\mathscr{A}}$. So, the equational perspective looks at how values of acceptance or rejection propagate within the attack graph stabilizing into steady states—the solutions—that have a nice correspondence with Dung's theory.

Example 4. Consider the graph on the left of Figure 1. The corresponding system of equations is:

$$\begin{aligned} f(a) &= 1 - \max(\{f(b)\}) \\ f(b) &= 1 - \max(\{f(a)\}) \end{aligned}$$

This gives three solutions: $f'(a) = 1$ and $f'(b) = 0$, $f''(a) = 0$ and $f''(b) = 1$, $f'''(a) = 0.5$ and $f'''(b) = 0.5$. The latter minimizes the set $\{a \mid f_g(a) = 1\}$ and corresponds therefore to the grounded extension.

6.2 Playing argument games through equations

We now look at how to build winning strategies for the proponent in an argument game using solutions to the system of equation of a given attack graph.

Let $\mathscr{A}$ be an attack graph. Consider its argument game $\mathscr{D}(\mathscr{A})@a$ for grounded at argument a and the equational theory for $\mathscr{A}$ corresponding to its grounded extension. Consider a strategy for $\mathscr{P}$ with the following property:

$$\sigma^*_{\mathscr{P}}(\mathbf{a}) \in \{a \mid f_g(a) = \max(\{b \mid b \in R(h(\mathbf{a}))\})\} \quad \text{FOR } \mathtt{t}(\mathbf{a}) = \mathscr{P} \qquad (10)$$

Intuitively, the strategy consist in $\mathscr{P}$ maximizing at each of her choice nodes the value f_g among the arguments attacking the last argument in the dialogue. In other words, $\mathscr{P}$ uses the information encoded by f_g to pick her arguments.

We can show that if the set of dialogues generated by $\sigma^*_{\mathscr{P}}$ are all of even length smaller than $2n$ then $\mathscr{P}$ can force a win in at most $2n$ rounds and vice versa:

Theorem 4 (Equationally defined winning strategies). *Let $\mathscr{D}(\mathscr{A})@a$ be the argument game for grounded on $\mathscr{A}$ instantiated at a, for $0 \leq n < \omega$:*

$$\forall \mathbf{a} \in T_{\sigma^*_{\mathscr{P}}} : \ell(\mathbf{a}) = 2m \leq 2n \Longleftrightarrow a \in Win^{2n}_{\mathscr{P}}$$

Proof (Sketch). [RIGHT TO LEFT] If $a \in Win^{2n}_{\mathscr{P}}$ then $\mathscr{P}$ can force a win in at most $2n$ rounds. By its definition (Formula (10)), $\sigma^*_{\mathscr{P}}$ must be a winning strategy. So, for any response $\sigma_{\mathscr{O}}$, $\mathtt{p}(\sigma^*_{\mathscr{P}}, \sigma_{\mathscr{O}}) = \mathscr{P}$ and hence the length $\ell(\sigma^*_{\mathscr{P}}, \sigma_{\mathscr{O}})$ must be even. Suppose now towards a contradiction that $\ell(\sigma^*_{\mathscr{P}}, \sigma_{\mathscr{O}}) > 2n$. $\mathscr{P}$ would then need in one case more than $2n$ rounds to win the game, against the assumption. [LEFT TO RIGHT] Similar.

In other words, $\sigma^*_{\mathscr{P}}$ is some kind of 'canonical' strategy for $\mathscr{P}$. As a direct corollary we obtain: $\sigma^*_{\mathscr{P}}$ is a winning strategy if and only if $f_g(a) = 1$. That is, a strategy that maximizes f_g at each choice node is winning for $\mathscr{P}$ if and only if the f_g value of the first argument is 1, i.e., if and only if a belongs to the grounded set. Similarly, it directly follows that if two arguments are strategically equivalent, then $\sigma^*_{\mathscr{P}}$ is winning (in a given number of rounds) for the first argument if and only if it is winning (in the same number of rounds) for the second.

6.3 Bisimulation, status equivalence, strategic equivalence and equational semantics

The equational semantics of abstract argumentation helps us in bringing together all the results handled in the paper, highlighting a wealth of interconnections between the modal, the strategic and the equational views of abstract argumentation theory.

Concretely, we have seen that frame bisimulation implies the status equivalence of two arguments in two attack graphs, which is in turn equivalent to their strategic equivalence in argument games seen as equivalence of 'powers' of strategies of the proponent and the opponent. All these different types of equivalences force arguments to obtain the same values in terms of Dung's semantics (i.e., one belongs to the grounded set if and only if the other also does) and Gabbay's equational variants (i.e., the value of f_g is the same for both arguments), as well as guaranteeing that equationally defined strategies for the proponent are winning on the first graph only if they are winning on the second, and vice versa. Figure 7 depicts these relations diagrammatically.

7 Conclusions

The paper has touched upon several strands of research at the interface of Dung-style abstract argumentation, modal logic, games and equational systems. From this interdisciplinary vantage point the paper has advocated a notion of equivalence of arguments abstracting from their content and based on the way they 'behave' with respect to the other arguments in the attack graph with respect to some external criterion of 'justifiability', which in this paper has been assumed to be the grounded extension.

First of all, the paper has shown how modal logic puts at disposal a number of notions and tools that can be readily used to provide an analysis of this sort of equivalence of arguments based on their abstract patterns of interaction. Argument

$$\begin{tikzcd}
\mathscr{A},a \rightleftarrows \mathscr{A}',a' \arrow[d, "\text{Fact 5}"'] & \sigma^*_{\mathscr{P}} \text{ wins in } \mathscr{A},a \Longleftrightarrow \sigma^*_{\mathscr{P}} \text{ wins in } \mathscr{A}',a' \arrow[d, leftrightarrow, "\text{Theorem 4}"'] \\
\mathscr{A},a \equiv_{\mathtt{d}} \mathscr{A}',a' \arrow[r, "\text{Fact 4}"] \arrow[d, leftrightarrow, "\text{Theorem 3}"'] & a \in \mathsf{lfpd}_{\mathscr{A}} \Longleftrightarrow a' \in \mathsf{lfpd}_{\mathscr{A}'} \arrow[d, leftrightarrow, "\text{Gabbay (2012)}"'] \\
\mathscr{A},a \text{ str. equiv. to } \mathscr{A}',a' & f_g(a) = f_g(a') = 1
\end{tikzcd}$$

Fig. 7 A diagram relating the notions of frame bisimulation, status equivalence, strategic equivalence, sameness of values according to Dung's grounded semantics, sameness of value according to Gabbay's equational semantics for the grounded set, and equivalence of 'powers' of equationally defined winning strategies

equivalence has been based on the notion of modal equivalence, and thereby related to the notion of (frame) bisimulation. This strengthens the many links between abstract argumentation and modal logic that have been object of several recent studies (e.g., (Caminada and Gabbay, 2009; Gabbay, 2011b) and (Grossi, 2009, 2010, 2011; Gratie et al, 2012)).

Second, the paper has shown how this static view of equivalence has a natural dynamic and strategic counterpart in argument games. In this view equivalent arguments are such that they support strategies for the proponent and opponent having the 'same powers' where power is intended as the possibility to guarantee a win in at most a given number of rounds of the game. This, together with the previous modal perspective, brings argumentation close to the thriving body of research into games and logical dynamics (van Benthem, 2011, 2013), and offers the picture of a theory that goes well beyond its more 'traditional' boundaries of the static study of justification criteria for arguments

Finally, Gabbay's equational approach (Gabbay, 2011a, 2012) to abstract argumentation has been used in relation to argument games as a method for constructing winning strategies for the proponent, thereby providing a sort of 'canonical' characterization of strategies viewed as local maximizers of the values provided by solutions to the equational systems of the graphs. This lays an interesting bridge between the modal and game-theoretic view of abstract argumentation and the rich body of techniques made available by the equational view.

Acknowldedgments

We would like to thank Johan van Benthem for the many useful inputs that helped us shape this last version of the paper.

References

Atkinson K, Bench-Capon T (2007) Argumentation and standards of proof. In: Proceedings of the 11th International Conference on Artificial Intelligence and Law (ICAIL'07), ACM, pp 107–116

Baroni P, Giacomin M (2009) Semantics of abstract argument systems. In: Rahwan I, Simari GR (eds) Argumentation in Artifical Intelligence, Springer

Baroni P, Caminada M, Giacomin M (2011) An introduction to argumentation semantics. The Knowledge Engineering Review 26(4):365–410

van Benthem J (1983) Modal Logic and Classical Logic. Monographs in Philosophical Logic and Formal Linguistics, Bibliopolis

van Benthem J (2002) Extensive games as process models. Journal of Logic, Language and Information 11:289–313

van Benthem J (2011) Logical Dynamics of Information and Interaction. Cambridge University Press

van Benthem J (2013) Logic in Games. MIT Press, in press

Blackburn P, de Rijke M, Venema Y (2001) Modal Logic. Cambridge University Press, Cambridge

Caminada M (2006) On the issue of reinstatement in argumentation. In: Fischer M, van der Hoek W, Konev B, Lisitsa A (eds) Logics in Artificial Intelligence. Proceedings of JELIA 2006, pp 111–123

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

Davey BA, Priestley HA (1990) Introduction to Lattices and Order. Cambridge University Press

Dung PM (1994) Logic programming as dialogue games. Tech. rep., Division of Computer Science, Asian Institute of Technology

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

Gabbay D (2011a) Introducing equational semantics for argumentation networks. In: Liu W (ed) Proceedings of ECSQARU 2011, no. 6717 in LNAI, pp 19–35

Gabbay D (2011b) Sampling logic and argumentation networks: A manifesto (volume 2). In: Gupta A, van Benthem J (eds) Logic and Philosophy Today, Studies in Logic, vol 30, College Publications, pp 231–250

Gabbay D (2012) An equational approach to argumentation networks. Argument and Computation 3(2–3):87–142

Gabbay D (2013) Meta-Logical Investigations in Argumentation Networks. College Publications, to appear

Gratie C, Florea AM, Meyer J (2012) Full hybrid mu-calculus, its bisimulation invariance and application to argumentation. In: Proceedings of COMMA 2012, pp 181–194

Grossi D (2009) Doing argumentation theory in modal logic. ILLC Prepublication Series PP-2009-24, Institute for Logic, Language and Computation

Grossi D (2010) On the logic of argumentation theory. In: van der Hoek W, Kaminka G, Lespérance Y, Sen S (eds) Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), IFAAMAS, pp 409–416

Grossi D (2011) Argumentation theory in the view of modal logic. In: McBurney P, Rahwan I (eds) Post-proceedings of the 7th International Workshop on Argumentation in Multi-Agent Systems, no. 6614 in LNAI, pp 190–208

Modgil S, Caminada M (2009) Proof theories and algorithms for abstract argumentation frameworks. In: Rahwan I, Simari G (eds) Argumentation in AI, Springer, pp 105–132

Osborne MJ, Rubinstein A (1994) A Course in Game Theory. MIT Press

Parikh R (2002) Social software. Synthese 132(3):187–211

Pollock JL (1987) Defeasible reasoning. Cognitive Science 11:481–518

Pollock JL (1991) A theory of defeasible reasoning. International Journal of Intelligent Systems 6(1):33–54

Prakken H (2009) Models of persuasion dialogue. In: Rahwan I, Simari G (eds) Argumentation in Artificial Intelligence, Springer, chap 14

Toulmin S (1958) The Uses of Argument. Cambridge University Press

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