Uncertainty in Abstract Argument Games: A Case Study on the Game for the Grounded Extension

Davide Grossi † *and Wiebe van der Hoek* ‡

University of Liverpool, UK
†d.grossi@liverpool.ac.uk
‡wiebe.van-der-hoek@liverpool.ac.uk

Abstract

The paper argues for the equipment of argument games with richer game-theoretic features. Concretely, it tackles the question of what happens to argument games when proponent and opponent are uncertain about the attack graph upon which they are playing. This simple sort of uncertainty, we argue, caters for the modeling of several strategic phenomena of real-life arguments. Using the argument game for the grounded semantics as a case study, the paper studies the impact of uncertainty over the ability of argument games to deliver adequacy with respect to their corresponding semantics.

1 Introduction

Abstract argumentation theory *à la* Dung offers a good level of abstraction from which to study the idea of "justifiability" of arguments, intended as abidance to a given standard of proof. Much research since [9] has been devoted to defining and studying such standards of proof—the so-called extensions—in terms of structural properties of attack graphs (see [19, 4, 2] for comprehensive overviews). Some research has then focused on operationalizations of those definitions via two-player games (see [19, 15] for extensive overviews).

The game-theoretic aspect of argumentation is evident in real-life argument, where strategizing is a vital component of the activities of disputing and debating. However, the sort of games studied thus far in argumentation are unable to capture many of the epistemic and strategic issues that seem to permeate real debates.

The present paper is part of a research program which, in the intention of the authors, should bring the theory of argument games to incorporate some of the rich mathematical toolbox of modern game theory. Specifically, the paper sets out to extend the standard theory of argument games to incorporate issues of uncertainty regarding the underlying attack graphs upone which the games are played. Many real-life arguments seem to have this feature, e.g.: two politicians debating in front of a heterogeneous audience consisting of groups of listeners evaluating the debate with respect to different attack graphs; or a plaintiff and a defence lawyer arguing in front of a judge whose interpretation of the available evidence is uncertain

Related work In recent years, techniques from theoretical economics have already found their way to abstract argumentation. However, interestingly enough, there has been a clear bias in favor of techniques coming from the field of social choice theory and mechanism design concerning, broadly speaking, the 'aggregation' of argumentation frameworks [6, 22, 20, 5].

Despite the game-like nature of real-life argumentation, applications of techniques coming from game theory proper have been much rarer. The abovementioned literature on the operationalization of Dung's semantics via games is in this sense the richest. Its focus are two-player games—called *argument games*—where the proponent of the game is assured to win the game if and only if the argument she claims at the beginning of the game satisfies a given standard of proof. This property is called adequacy and constitutes also the focus of the present paper. Concretely, we will be aiming at generalizing the notion of adequacy to cover games in which proponent and opponent may be uncertain about the graph upon which they are playing the argument game.

Besides the abovementioned literature, we are only aware of [14], which studies a strategic zero-sum game with randomization, using it to obtain quantitative refinements of some of Dung's semantics. Recently, a few game theorists have also taken up the challenge of modeling some features of real-life debates (e.g., [11]) but those contributions are framed at an even higher level of abstraction than the one enabled by Dung's framework. Neither [14] nor [11] are interested in the question of adequacy of a game with respect to a given semantics.

Structure of the paper We will start by introducing just as much of abstract argumentation that will be needed in the paper, and in particular the notion of grounded extension. We then proceed by discussing in detail a game for the grounded extension and proving its adequacy. That game is the main building block of the present paper. We then introduce an original notion of uncertainty for abstract argumentation consisting of a probability distribution over a set of attack graphs. This notion of uncertainty is then studied providing a probabilistic version of the grounded extension and two argument games modeling how proponent and opponent can cope with that type of uncertainty. Issues of adequacy concerning these games are then discussed, and a brief discussion section concludes the paper.

2 Abstract argumentation

2.1 Attack graphs

We start by introducing the notion of argumentation framework, due to [9], which we call here attack graph:

Definition 1 (Attack graphs)**.** *An attack graph is a tuple $\mathcal{A} = \langle A, \rightarrow \rangle$ where:*

- *A is a finite 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 for $a \rightarrow X$. Given an argument a, we denote by $[\![a]\!]_{\mathcal{A}}$ the set of arguments attacking a: $\{b \in A \mid b \rightarrow a\}$.

These relational structures are the building blocks of abstract argumentation theory. The set A represents a set of further unanalyzed arguments or pieces of evidence.[1] Relation $\rightarrow$, called the 'attack' relation, encodes the way arguments attack one another. So, an attack graph is best viewed as a high-level representation of the conflicts inherent within the information put forth by a set of arguments or, also, as the atemporal representation of a debate where all evidence has been revealed.

2.2 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*. The study of structural properties of attack graphs provides therefore very general insights on how competing arguments interact and how collections of them form 'tenable' or 'justifiable' argumentative positions.

2.2.1 The grounded extension

While many different structural standards of proof have been defined and studied since [9], in this paper we focus on just one of them, the so-called grounded extension. This choice is dictated by properties of the grounded extension—namely its uniqueness—that make it a particularly elegant concept, as well as by the fact that the game corresponding to the grounded extension—to which we will turn in the

[1] Bearing the abstract nature of attack graphs in mind, we will often feel free to interpret the elements of A both as arguments or evidences, although we will consistently use the term 'argument' throughout the paper.

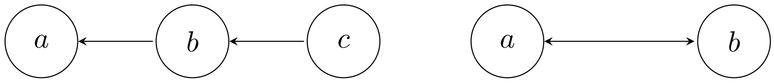

Fig. 1: Two simple examples of attack graphs.

next section—has been the first argument game studied in the literature [8] and is, arguably, the one that is best understood.

The grounded extension is defined via this function:

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

$$\mathtt{d}_{\mathcal{A}}(X) \quad = \quad \{a \in A \mid \forall b \in A : \text{ IF } b \rightarrow a \text{ THEN } X \rightarrow b\}$$

For $0 \leq k < \omega$ we denote by $\mathtt{d}^k_{\mathcal{A}}(X)$ the k^{th} iteration of function $\mathtt{d}_{\mathcal{A}}$ on set X.

The function outputs, for each set of arguments, the set of arguments defended by X. A set of argument X such that $X \subseteq \mathtt{d}_{\mathcal{A}}(X)$ is therefore able to defend itself from external attacks. Intuitively, function $\mathtt{d}$ encodes how much each set of arguments is able to defend in a given graph.

Definition 3 (The grounded extension)**.** *Let $\mathcal{A} = \langle A, \rightarrow \rangle$ be an attack graph. The grounded extension of $\mathcal{A}$ is the least fixpoint of $\mathtt{d}_{\mathcal{A}}$, in symbols: $\mathsf{lfp}(\mathtt{d}_{\mathcal{A}})$.*

In other words, an argument is grounded if and only if it belongs to the smallest fixpoint of the defence function. So, intuitively, it represents the smallest set of arguments that support themselves. As such, it can be viewed as a very demanding standard of proof. It is finally worth recalling that as a direct consequence of the Knaster-Tarski theorem[2] and the fact that the characteristic function can be easily shown to be monotonic, the least fixpoint of the characteristic function exists for any attack graph and equals the intersection of all pre-fixpoints of the defence function, i.e., $\bigcap \{X \subseteq A \mid \mathtt{d}_{\mathcal{A}}(X) \subseteq X\}$.

Example 1. *In Figure 1 we have:*

- *In the graph on the left: $\mathtt{d}(\emptyset) = \{c\}$, $\mathtt{d}(\{c\}) = \mathtt{d}^2(\emptyset) = \{a, c\}$ and $\mathtt{d}(\{a, c\}) = \mathtt{d}^3(\emptyset) = \{a, c\}$, which is the grounded extension.*
- *In the graph on the right: $\mathtt{d}(\emptyset) = \emptyset$, which is the grounded extension.*

[2] See [7].

3 Argument games

While the study of different formal definitions of extensions constitutes the main body of abstract argumentation theory, many researchers in the last two decades have focused on 'proof procedures' for argumentation, i.e., procedures able to adequately establish whether a given argument belongs or not to a given extension. Many of such proof procedures have resorted to abstractions coming from game theory and have given rise to a number of different types of games, called *dialogue* or *argument* games.[3]

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.

3.1 An argument game for the grounded extension

This section introduces an argument game for the grounded extension (*game for grounded* in short), which will be the starting point of the investigations presented here. Although quite some literature on variants of this game already exists, we feel the need to furnish this introductory section with a fair amount of detail.

We do this for several reasons: first, different variants of the game have been proposed (e.g., in [19, 15]); second, the presentations of the game in the literature are not always in line with formats of presentation typically followed in game theory, which is a drawback for the purpose of the present paper, specifically aiming at a closer link with game theory; third, the only published[4] proof of adequacy of a game for grounded we are aware of, namely the one given in [15], adopts an unusual—with respect to adequacy proofs for games in, e.g., logic or theoretical computer science—format where winning strategies are subjected to extra conditions[5]. The proof we will give is of the standard form mentioned above, i.e., "argument a belongs to the grounded extension if and only if the proponent has a winning strategy in the game for the grounded extension started at a". Casting adequacy in this format is an essential stepping stone for the game-theoretic study of these games we pursue in the rest of the paper and its probabilistic generalization (Theorem 2).

[3] The contributions that started this line of research are the unpublished technical report [8] and [13, 23]. Cf. [15] for a recent overview.

[4] According to [19], a proof was given in the unpublished technical report [8].

[5] Roughly, the format adopted there reads: "a belongs to the grounded extension if and only if the proponent has a winning strategy *such that the set of proponent's arguments in the winning strategy is conflict free*".

Dialogues	$\mathcal{P}$ wins	$\mathcal{O}$ wins
$\ell(\mathbf{a}) < \omega$	$\mathtt{t}(s) = \mathcal{O}$	$\mathtt{t}(s) = \mathcal{P}$
$\ell(\mathbf{a}) = \omega$	*never*	*always*

Tab. 1: Winning conditions for the game for grounded given a terminal dialogue $\mathbf{a}$.

3.1.1 Notation

We will need the following notation. 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 of $\mathbf{a}$, 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 if $\mathbf{a}|_n = \mathbf{a}$, and ω otherwise.

3.1.2 The game

We start with the formal definition, but the reader might wish to combine this with the informal reading coming next.

Definition 4 (Argument game for grounded)**.** *The argument game for grounded is a function $\mathcal{D}(.)$ which for each attack graph $\mathcal{A}$ yields structure $\mathcal{D}(\mathcal{A}) = \langle N, A, \mathtt{t}, \mathtt{m}, \mathtt{p} \rangle$ where:*

- *$N := \{\mathcal{P}, \mathcal{O}\}$—the set of players consists of proponent $\mathcal{P}$ and opponent $\mathcal{O}$.*
- *A is the set of arguments in $\mathcal{A}$.*
- *$\mathtt{t} : A^{<\omega} \longrightarrow N$ is the* turn function. *It is a (partial[6]) 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 $\mathtt{t}(\mathbf{a}) = \mathcal{P}$, and if $\ell(\mathbf{a}) = 2m + 1$ then $\mathtt{t}(\mathbf{a}) = \mathcal{O}$. I.e., even positions are assigned to the proponent and odd positions to the opponent.*
- *$\mathtt{m} : A^{<\omega} \longrightarrow A$ is a (partial) function defined as:*

$$\mathtt{m}(\mathbf{a}) = [\![h(\mathbf{a})]\!]_{\mathcal{A}}$$

  I.e., the accessible 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.

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

- $\mathtt{p} : T \longrightarrow N$ *is the payoff function in Table 1 associating a player—the winner—to each terminal dialogue.*

Dialogue games are played starting from a given argument a. When a is explicitly given we talk about an instantiated dialogue game (notation, $\mathcal{D}(\mathcal{A})@a$).

Here is an informal description of the game. The two players play the game by alternating each other (opponent starts) and navigating the attack graph along the 'being attacked' relation, both having the same move function. The winning conditions state that the proponent wins whenever she manages to state an argument to which the opponent cannot reply, i.e., an argument with no attackers. Then, in accordance with the saying 'he who laughs last laughs best', she wins. However, notice that the winning conditions are somewhat asymmetric as the proponent not only loses when she is stuck with no counter-arguments, but also when the game loops in an infinite dispute.[7]

Remark 1 (Argument games & game theory)**.** *From the point of view of game theory,[8] the games in Definition 4 can be classified as two-player zero-sum (the range of payofss being $\{1, 0\}$) extensive games with perfect information and with fully aware players. In argumentation-theoretic terms, this means that those games model arguments with the following properties: two arguers exchange arguments; of the two arguers one wins the debate, while the other loses; both arguers always know what the current argument is, i.e., intuitively, they always know what position they occupy in the game with respect to the underlying attack graph.*

3.1.3 Strategies and winning positions

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

Definition 5 (Strategies)**.** *Let $\mathcal{D} = \langle N, A, \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$, $a \in A$ and $i \in N$. A strategy for i in the instantiated game $\mathcal{D}@a$ is a function: $\sigma_i : \{\mathbf{a} \in D - T \mid \mathbf{a}_0 = a \text{ AND } \mathtt{t}(\mathbf{a}) = i\} \longrightarrow A$ telling i which argument to chose at each non-terminal dialogue $\mathbf{a}$ in $\mathcal{D}@a$. The set of terminal dialogues compatible with σ_i is defined as follows: $T_{\sigma_i} = \{\mathbf{a} \in T \mid \text{ IF } \mathtt{t}(\mathbf{a}|_n) = i \text{ THEN } \mathbf{a}_{n+1} = \sigma_i(\mathbf{a}|_n)\}$.*

So, in the game for grounded, a strategy $\sigma_{\mathcal{P}}$ will encode the proponent's choices in dialogues of odd length, while $\sigma_{\mathcal{O}}$ will encode the opponent's choices in dialogues of even length. Observe that, in a game for grounded, a strategy $\sigma_{\mathcal{P}}$ and a strategy $\sigma_{\mathcal{O}}$—i.e., a strategy profile in the game-theory terminology—together determine one terminal dialogue or, in other words, $T_{\sigma_i \mathcal{P}} \cap T_{\sigma_{\mathcal{O}}}$ 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, and the related one of winning position.

[7] Infinite arguments might sound puzzling, but they should simply be taken as models of an irresolvable deadlock, e.g.: "yes it is!", "no it isn't!", "yes it is!", "no it isn't!".

[8] Cf. [16] for an overview.

Definition 6 (Winning strategies and arguments)**.** *Let $\mathcal{D} = \langle N, S, \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$, $a \in A$ and $i \in N$.*

- *A strategy σ is* winning for *i in $\mathcal{D}@a$ 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 $\mathcal{D}@a$. The set of winning positions of $\mathcal{D}$ for i is denoted $Win_i(\mathcal{D})$.*
- *An argument a is* winning for *i* in *k* rounds *iff there exists a winning strategy for i in $\mathcal{D}@a$ 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 steps. The set of winning positions in k rounds is denoted $Win^k_i(\mathcal{D})$.*

As observed in Remark 1 dialogue games are two-player zero-sum games with perfect information. It follows from the so-called Zermelo's theorem [24] that these games are determined, in the sense that either $\mathcal{P}$ or $\mathcal{O}$ possesses a winning strategy, and hence that each argument in an attack graph is either a winning position for $\mathcal{P}$ or a winning position for $\mathcal{O}$.

Example 2. *Consider Figure 2 with the following interpretation of the attack graph: a = "The witness saw the defendant leaving the crime scene a few minutes after the murder happened", b = "The witness is unreliable as she is short-sighted", c = "Although short-sighted, the witness is reliable since she was waring glasses", d = "The witness is unreliable as the crime scene was not illuminated and she was too far away to recognize the defendant". Here $\mathcal{P}$ has no winning strategy, but not because of any infinite dispute, rather, simply because $\mathcal{O}$, at his first turn, can move to argument d—which is his winning strategy—where $\mathcal{P}$ cannot reply. Note that if $\mathcal{O}$ makes a 'mistake' by playing b then $\mathcal{P}$ can win by reinstating a via c.*

3.1.4 Adequacy

Now all ingredients are in place to study the property we are interested in, viz. the adequacy of the game of Definition 4 with respect to the grounded extension. We first prove a slightly stronger result: an argument a belongs to the k^{th} iteration of the characteristic function on the empty set of arguments, if and only if $\mathcal{P}$ has a winning strategy in the game initiated at a, which she can carry out in at most $2k$ steps.

Lemma 1 (Strong adequacy of the game for grounded)**.** *Let $\mathcal{D}(\mathcal{A}) = \langle N, S, \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$ be the dialogue game for grounded on graph $\mathcal{A}$ and $a \in A$:*

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

Proof. We proceed by induction on the depth d of the subtree $\sigma(\mathcal{D}(\mathcal{A})@a)$ yielded by the winning strategy σ (recall Definition 5).

[B:] We have the following equivalences:

$$\begin{array}{rcll} a \in \mathtt{d}^0_{\mathcal{A}}(\emptyset) & \Longleftrightarrow & \nexists b : b \rightarrow a & \text{[Definition 2]} \\ & \Longleftrightarrow & a \in \mathit{Win}^0_{\mathcal{P}}(\mathcal{D}(\mathcal{A})) & \text{[Definition 6]} \end{array}$$

[S:] If $a \in \mathtt{d}^n_{\mathcal{A}}(\emptyset) \Longleftrightarrow a \in \mathit{Win}^{2n}_{\mathcal{P}}(\mathcal{D}(\mathcal{A}))$ (IH) then $a \in \mathtt{d}^{n+1}_{\mathcal{A}}(\emptyset) \Longleftrightarrow a \in \mathit{Win}^{2(n+1)}_{\mathcal{P}}(\mathcal{D}(\mathcal{A}))$. [LEFT TO RIGHT] Assume $a \in \mathtt{d}^{n+1}_{\mathcal{A}}(\emptyset)$. This means that $\forall b : b \rightarrow a, \exists c : c \rightarrow b$ and such that $c \in \mathtt{d}^n_{\mathcal{A}}(\emptyset)$ which, by IH, is equivalent to $c \in \mathit{Win}^{2n}_{\mathcal{P}}(\mathcal{D}(\mathcal{A}))$. So, by Definition 4, for any $\mathcal{O}$'s move b at position a, $\mathcal{P}$ has a counter-argument c from which she has a winning strategy in $2n$ rounds. Hence, by Definition 6, $\mathcal{P}$ can win the game at a in $2n+2$ rounds, i.e., $a \in \mathit{Win}^{2(n+1)}_{\mathcal{P}}(\mathcal{D}(\mathcal{A}))$. [RIGHT TO LEFT] Assume $a \in \mathit{Win}^{2(n+1)}_{\mathcal{P}}(\mathcal{D}(\mathcal{A}))$. This means that, for any $\mathcal{O}$'s move b at a, $\mathcal{P}$ has a counter-argument c from which she has a winning strategy in $2n$ rounds. By IH, this is equivalent with $c \in \mathtt{d}^n_{\mathcal{A}}(\emptyset)$ and by Definition 2 we conclude that $a \in \mathtt{d}^{n+1}_{\mathcal{A}}(\emptyset)$. 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.

Theorem 1 (Adequacy of the game for grounded)**.** *Let* $\mathcal{D}(\mathcal{A}) = \langle N, S, \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$ *be the dialogue game for grounded on graph* $\mathcal{A}$ *and* $a \in A$*:*

$$a \in \mathsf{lfp}(\mathtt{d}_{\mathcal{A}}) \Longleftrightarrow a \in \mathit{Win}_{\mathcal{P}}(\mathcal{D}(\mathcal{A})).$$

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

$$\begin{array}{rcl} a \in \mathit{Win}_{\mathcal{P}}(\mathcal{D}(\mathcal{A})) & \Longleftrightarrow & a \in \bigcup_{0 \leq k < \omega} \mathit{Win}^{2k}_{\mathcal{P}}(\mathcal{D}(\mathcal{A})) \\ & \Longleftrightarrow & a \in \bigcup_{0 \leq k < \omega} \mathtt{d}^k_{\mathcal{A}}(\emptyset) \\ & \Longleftrightarrow & a \in \mathsf{lfp}.\mathtt{d}_{\mathcal{A}} \end{array}$$

The first equivalence holds by the winning conditions of Definition 4 and Definition 6: $\mathcal{P}$ 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 1 and the third one by the finiteness assumption in Definition 1 and general fixpoint theory.[9] □

3.2 On the game-theoretic 'meaning' of adequacy

The study of argument games has been thrust by their use as proof procedures for Dung's semantics [15]. In fact, the importance of theorems like Theorem 1 is

[9] See [7, Ch. 8].

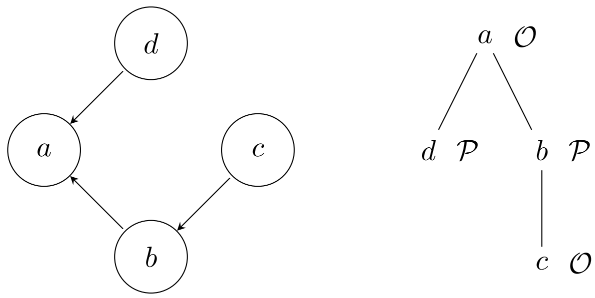

Fig. 2: An attack graph (left) and its dialogue game for grounded (right). Positions are labeled by the player whose turn is to play. $\mathcal{P}$ wins the terminal dialogue abc but loses the terminal dialogue ad.

that they guarantee the argument game at issue to be a sound (if proponent has a winning strategy then the the argument is grounded) and complete (if the argument is grounded, then proponent has a winning strategy) proof procedure with respect to the corresponding semantics.

However, beyond their mere proof-theoretic use, literature in argumentation has also advanced a view of these games as viable models of procedural rules, or protocols, for debates (good examples of this line of research are [18, 17]). At the same time, other literature (e.g., [1]) has pointed out, convincingly in our view, that Dung's extensions can be soundly viewed as abstract models of standards of proof in debates. Viewed in this light, adequacy has rather to do with whether a given debate protocol is successful in *implementing* a given standard of proof. We use the word "implement" here in the technical sense in which it is typically used in game theory[10] or social software [21].

Elaborating on this, a standard of proof like grounded can be viewed as a social choice function deciding for each argument whether the proponent should be considered to have proven the argument, or whether the opponent should be considered to have disproven it. The game will then let proponent and opponent interact in a strategic setting and will be said to implement the standard of proof—or to be adequate with respect to it—whenever the outcome dictated by the standard of proof coincides with the outcome reached by the game in equilibrium, where equilibrium means, in the limited context of zero-sum games (recall Remark 1), the outcome that either of the players can enforce by playing her winning strategy.

The remaining of the paper will push this game-theoretic view of adequacy, investigating how the ability of argument games to successfully implement standards of proof fares in presence of uncertainty.

[10] We refer in particular to the so-called implementation theory. See [16, Ch. 10].

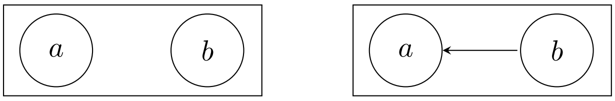

Fig. 3: The belief space of Example 3. We will refer to the graph on the left as $\mathcal{A}_L$ and the one on the right as $\mathcal{A}_R$.

4 Uncertainty in argumentation

In this section we look at structures that model a type of uncertainty concerning the lack of information about what the attack graph upon which a given standard of proof—in our case the grounded extension—should be applied.

4.1 Beliefs

Let $\mathbf{A}$ be a finite set of attack graphs over a given set of arguments A: $\mathbf{A} \subseteq \mathfrak{A}(A)$. A belief on $\mathbf{A}$ is a probability distribution $\Delta(\mathbf{A})$ over A (the belief space). A belief will be denoted as a sequence $\alpha_1, \ldots, \alpha_n$ of probabilities $0 < \alpha \leq 1$ summing up to 1, e.g.: $B = \langle 0.5, 0.5 \rangle$ or $B = \langle 0.1, 0.8, 0.1 \rangle$. We will take care that the order of the sequence provides the information to assign the right probability to the right graph.

It is worth stressing that—as customary in game theory —these probabilities can be legitimately read in several different ways, all getting interesting interpretations in the context of argumentation. We mention the two we will be using in the paper: first, a probability distribution can be interpreted as the belief that one agent holds when uncertain about the actual attack graph upon which a given argument will be evaluated, e.g., the attack graph endorsed by a listener; second, a probability distribution can be interpreted as the proportions of different 'types' of listeners in a population of listeners, viz. the percentage of agents endorsing graph $\mathcal{A}_1$, the percentage of agents endorsing graph $\mathcal{A}_2$, and so on. We illustrate these notions with two toy examples:

Example 3. *Two politicians have to openly debate during the election campaign about cutting taxes vs. increasing public services: let argument* a *be "taxes should be cut" and* b *be "public services should be improved". They both believe that* 70% *of the public do not consider the two arguments incompatible and are ready to support both, while the remaining* 30% *do consider the two arguments incompatible and, moreover, find argument* b *stronger than* a *(see Figure 3). The politicians' belief are modeled by distribution* $B = \langle 0.7, 0.3 \rangle$.

Example 4. *Assume again the arguments* a, b, c *and* d *of Example 2. A prosecutor and a defense lawyer are arguing about argument* a*. They are uncertain of how the judge is going to interpret the evidence at hand. They believe that it is* 10%

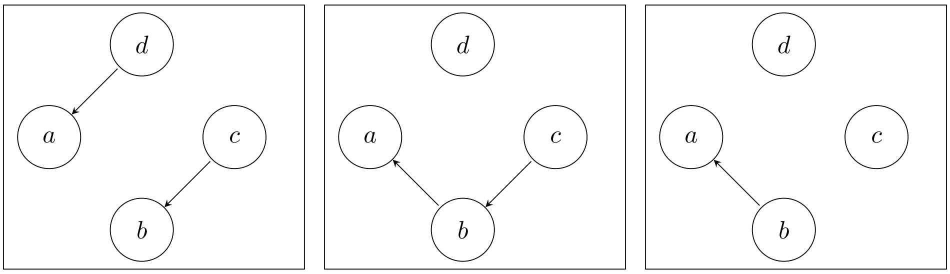

Fig. 4: The belief space of Example 4. In further discussing the example, we will refer to the three graphs as $\mathcal{A}_1, \mathcal{A}_2$ and $\mathcal{A}_3$ in the obvious order.

probable that the judge would not recognize b as legitimate evidence against a, while accepting the other the evidence of d against a and c against b. They also believe that with probability 50% the judge will not accept evidence d against a, while admitting c against b and b against a. Finally, they believe with probability 40% that she would admit b against a but reject d against a and of c against b. This belief space is represented in Figure 4.1 and the belief is $B = \langle 0.1, 0.5, 0.4 \rangle$.

4.2 Probability of being grounded

A belief induces, for each argument, the probability of it to be proven with respect to the fixed standard of proof. Let $g : \mathbf{A} \times A \longrightarrow \{0, 1\}$ be the function recording whether a given argument belongs to the grounded extension of a given graph in the belief space $\mathbf{A}$:

$$g(\mathcal{A})(a) = \begin{cases} 1 & \text{IF } a \in \mathsf{lfp}(\mathsf{d}_{\mathcal{A}}) \\ 0 & \text{OTHERWISE} \end{cases} \tag{1}$$

With this auxiliary function at hand, we can formulate the following definition:

Definition 7 (Grounded in belief)**.** *Let $B = \Delta(\mathbf{A})$ be a belief and a an argument. The probability of a to belong to a grounded extension given B is given by the following formula:*

$$p_B(a) = \sum_{\mathcal{A} \in \mathbf{A}} B(\mathcal{A}) \cdot g(\mathcal{A})(a).$$

So, a is said to be grounded in B with probability α if and only if $p_B(a) = \alpha$.

Obviously, by Theorem 1, value $p_B(a)$ also expresses the probability of the proponent to have a winning strategy in the argument game for grounded played at a given belief B. Similarly, $1 - p_B(a)$ denotes then the probability that a does not belong to a grounded extension, and therefore the probability that the opponent has a winning strategy in the argument game for grounded instantiated at a.

Example 5 (Example 3 continued)**.** *Recall Example 3. We have quite simply that $p_B(a) = 0.7$ and $p_B(b) = 1$.*

Remark 2 (Audiences). *Examples 3 and 4 have illustrated the idea of a probability distribution over attack graphs in terms of proportions of types of listeners, or as plain uncertainty over the type of one listener. The role of listeners or audiences has already been object of attention within the argumentation theory community (e.g., [12, 3]). In those works audiences are modelled as extensions of attack graphs inducing preferences or rankings over sets of arguments, which then refine the evaluation of arguments in possibly different ways. In essence, we pursue the very same aim here, namely the representation of different ways of evaluating arguments. However, in order to more easily link with game-theoretic concepts, we do it via probability distributions over graphs rather than by introducing extra preferential structure within attack graphs.*

5 Arguing on the basis of beliefs

The previous section has presented probability distributions over attack graphs—which we called beliefs—as a viable modeling tool for a type of uncertainty in abstract argumentation, and it has shown how it naturally yields a probabilistic version of the grounded extension. This section addresses the related question of how an argument game would proceed under the assumption that proponent and opponent hold a same belief.

We contemplate two possibilities which, from a game-theoretic point of view, lead to two very different models:

Perfect information: In this case the belief is taken to encode the expectation of an agent with respect to a lottery drawn on the possible attack graphs, under the assumption that the result of the lottery can be observed before proponent and opponent start exchanging arguments. In other words, a move by Chance selects an attack graph, with the probabilities dictated by the belief, after which proponent and opponent play the game for grounded on the selected graph.

Imperfect information: In this case the belief is taken to encode the expectation of an agent with respect to the same lottery, but under the assumption that the result of the lottery cannot be observed before the argumentation starts. Like above, a move by Chance first selects an attack graph, but proponent and opponent do not know which one has been selected.

The next sections study the sort of games arising by these two interpretations of the uncertainty represented in a belief, with a particular emphasis on how they influence adequacy with respect to the grounded semantics.

5.1 Perfect information on the initial chance move

Example 6 (Example 3 continued). *During the campaign the two politicians will be invited by one journalist to a face off debate about the issues a and b. The* *pool of possible journalists interviewing them is representative of the* $\langle 0.7, 0.3\rangle$ *distribution in the public opinion, and both politicians know the 'type' of each journalist, i.e., whether she endorses the graph on the left of Figure 3 or the one on the right. Before the inverview even takes place the politician arguing for* a *can thus be sure she will argue successfully—she will convince the journalist—in at least* 70% *of the cases.*

Abstractly, proponent and opponent do not know which attack graph Chance will select, but they know that each graph will be selected with a given probability and that they will be able to observe it upon starting to play.

5.1.1 The game

The key idea behind the game is that the probability distribution on the underlying graphs determine the payoff a player obtains when the opponent runs out of arguments on the selected graph.

Definition 8 (Argument games with perfect information on Chance moves)**.** *The perfect information argument game is a function* $\mathcal{D}^P(.)$ *which for each belief* B *yields structure* $\mathcal{D}^P(B) = \langle N^P, A^P, \mathtt{t}^P, \mathtt{m}^P, \mathtt{p}^P\rangle$ *where:*

- $N^P = \{\mathcal{P}, \mathcal{O}, \mathit{Chance}\}$.
- $A^P = \{\langle a, \mathcal{A}\rangle \mid a \in A \text{ AND } \mathcal{A} \in \mathbf{A}\} \cup \{\iota\}$, *i.e., the set of positions of the game are arguments indexed by a graph, plus an initial 'empty' argument* ι*. Dialogues are denoted* $\mathbf{s}$*. Given a pair* $\langle a, \mathcal{A}\rangle$*, notation* $\langle a, \mathcal{A}\rangle_l$ *denotes* a*, viz. the argument in the pair, and* $\langle a, \mathcal{A}\rangle_r$ *denotes* $\mathcal{A}$*, viz. the graph in the pair. Similarly, for a dialogue* $\mathbf{s}$*, we denote by* $\mathbf{s}_l$ *the sequence of arguments occurring in the pairs in* $\mathbf{s}$*, i.e., if* $\mathbf{s}$ *is* $\langle a_1, \mathcal{A}_1\rangle, \ldots, \langle a_n, \mathcal{A}_n\rangle, \ldots$ *then* $\mathbf{s}_l$ *is* $a_1, \ldots, a_n, \ldots$*.*
- $\mathtt{t}^P$ *is defined exactly like* $\mathtt{t}$ *in Definition* 4 *except for the fact that* $\mathtt{t}^P(\iota) = \mathit{Chance}$*. In dialogues starting with* ι*, to keep the turn of* $\mathcal{O}$ *and* $\mathcal{P}$ *at even and, respectively, odd positions with respect to the initial argument chosen by Chance, we use a length function* ℓ^* *so defined:* $\ell^*(\mathbf{s}) = \ell(\mathbf{s}_l) - 1$ *for* $\mathbf{a}_1 = \iota$*.*
- $\mathtt{m}^P$ *modifies* $\mathtt{m}$ *of Definition 4 as follows. For* $\mathbf{s} = \langle\iota\rangle$, $\mathtt{m}^P(\mathbf{s}) \in \{\{\langle a, \mathcal{A}\rangle \mid \mathcal{A} \in \mathbf{A}\}\}_{a\in A}$*, i.e., Chance chooses among sets of argument-graph pairs, where the argument is kept stable. Otherwise, if* $\mathbf{s} \neq \langle\iota\rangle$, $\mathtt{m}^P(\mathbf{s}) = \{\langle a, \mathcal{A}'\rangle \mid h(\mathbf{s}_l) \leftarrow a \text{ AND } \mathcal{A}' = h(\mathbf{s}_r)\}$*. I.e., at a dialogue ending with* $\langle a, \mathcal{A}\rangle$*, only counterarguments in* $\mathcal{A}$ *may be selected.*[11] *The set of terminal dialogues* T^P *is defined in the obvious way.*

[11] Notice that, as a consequence, for all pairs $\langle a, \mathcal{A}\rangle$, $\langle a', \mathcal{A}'\rangle$ in a dialogue compliant with $\mathtt{m}^P$, $\mathcal{A} = \mathcal{A}'$.

- $\mathtt{p}^P : \{\mathcal{P}, \mathcal{O}\} \times T^P \longrightarrow [0, 1]$ *is the payoff function defined as follows:*

$$\mathtt{p}^P(i)(\mathbf{s}) = B(\mathcal{A}) \cdot w(i)(\mathbf{s}_l)$$

where $\mathcal{A}$ is the graph occurring in all pairs in $\mathbf{s}$ and $w(i)(\mathbf{s}) = \begin{cases} 1 & \text{IF } \mathtt{p}(\mathbf{s}_l) = i \\ 0 & \text{OTHERWISE} \end{cases}$ *encodes for a given terminal dialogue $\mathbf{s}$ in and player i, whether i wins the dialogue $\mathbf{s}_l$ in the underlying game for grounded. I.e., for i, the payoff function assigns to each terminal dialogue $1 \cdot B(\mathcal{A})$ if that dialogue is winning for i (according to the game for grounded) in the underlying graph, and 0 otherwise.*

This game is always initiated at the designated empty argument ι. We say that $\mathcal{D}^P(B)$ is instantiated at argument a (in symbols, $\mathcal{D}^P(B)@a$) if and only if $\mathtt{m}(\iota) = \{\langle a, \mathcal{A} \mid \mathcal{A} \in \mathbf{A}\rangle\}$, i.e., if Chance can choose only initial moves containing argument a.

Figure 5 illustrates the game as it models the perfect information interpretation of Examples 3 and 4. More concisely, the game can also be viewed as a probability distribution over the set of argument games based on the graphs in the belief space. This distribution represents the moves available to Chance and their respective probability and is taken, in the definition of the payoff function, to weight the wins and losses of proponent and opponent.

Remark 3. *A few characteristics of the game must be observed: first, by the set up of the payoff function, a player will obtain a non-negative payoff in a terminal dialogue if and only if the opponent has a payoff 0; second, the oucome of the game can rightly be seen as a probability distribution over the simple state space $\{\mathcal{P}\ wins, \mathcal{P}\ loses\}$ where the initial move by Chance is the randomizing device.*

5.1.2 Strategies

The definitions of strategy, winning strategy and winning argument are modified to account for the new structure of the game.

Definition 9. *Let $\mathcal{D}^P(B)@a$ be a perfect information game for belief B instantiated at a, and $i \in \{\mathcal{P}, \mathcal{O}\}$:*

- *A strategy for i is a tuple $\rho = \langle \sigma_i^{\mathcal{A}} \rangle_{\mathcal{A} \in \mathbf{A}}$ where each $\sigma_i^{\mathcal{A}}$ is a strategy of i in $\mathcal{D}(\mathcal{A})@a$, viz., the game for grounded on $\mathcal{A}$.*[12]
- *The value of strategy ρ_i is determined by the sum of the minimal payoffs determined by each strategy. Let $v(\sigma_i^{\mathcal{A}}) = \min \left\{ \mathtt{p}^P(i)(\mathbf{s}) \mid \mathbf{s}_l \in T_{\sigma_i^{\mathcal{A}}} \right\}$ be*

[12] To be precise ρ should assign pairs $\langle a, \mathcal{A} \rangle$ at each choice point, but since the graph remains constant throughout any dialogue, our formulation is considerably simpler.

the minimal payoff obtainable by i playing σ. The value of ρ_i is then defined by $\sum_{\sigma_i^{\mathcal{A}} \in \rho_i} v(\sigma_i^{\mathcal{A}})$.[13]

- *The set of arguments for which i has a strategy with value* at most α *is denoted $Win_i^{\alpha}(\mathcal{D}^P(B))$.*

So, the way proponent and opponent can play this game is by randomizing, according to their belief, over winning strategies they have available in the underlying games for grounded. Each winning strategy in the underlying games will contribute a strictly positive value to their payoff.

5.1.3 Probabilistic adequacy

The game presented can be shown to be adequate with respect to the probabilistic version of the grounded extension, yielding a probabilistic generalization of the original adequacy result.[14]

Theorem 2 (Probabilistic adequacy)**.** *Let $\mathcal{D}^P(B)$ be a perfect information game for belief B, $a \in A$ and $0 < \alpha \leq 1$:*

$$\alpha = p_B(a) \iff a \in Win_{\mathcal{P}}^{\alpha}(\mathcal{D}^P(B))$$

Sketch of proof. Let $\mathbf{A}_a = \{\mathcal{A} \in \mathbf{A} \mid g(\mathcal{A})(a) = 1\}$. From Definition 7 we obtain the following equations:

$$p_B(a) = \sum_{\mathcal{A} \in \mathbf{A}} B(\mathcal{A}) \cdot g(\mathcal{A})(a) = \sum_{\mathcal{A} \in \mathbf{A}_a} B(\mathcal{A})$$

We prove the implication in both directions. [LEFT TO RIGHT] Assume $\alpha \leq p_B(a)$. By Theorem 1 it follows that there exists a largest set of strategies $\left\{\sigma_{\mathcal{P}}^{\mathcal{A}_i}\right\}_{1 \leq i \leq n}$ where $n = |\mathbf{A}|$, such that each $\sigma_{\mathcal{P}}^{\mathcal{A}_i}$ is winning in $\mathcal{D}(\mathcal{A}_i)@a$. From which we have that $\alpha = \sum_{\mathcal{A} \in \mathbf{A}} v(\sigma_i^{\mathcal{A}})$ and hence, by Definition 9, $a \in Win_{\mathcal{P}}^{\alpha}(\mathcal{D}^P(B))$. [RIGHT TO LEFT] This direction is similar and can be obtained by arguing again by Theorem 1 and Definition 9. □

Intuitively, an argument belongs to the grounded extension with a given probability if and only if that probability is the best payoff that proponent can guarantee for herself in the game. The game can therefore be viewed as a procedure successfully implementing the standard of proof modeled by the grounded extension in settings where there is a shared uncertainty between proponent and opponent.

[13] Since the game at issue is strictly competitive, we are assuming that players play by maximizing their minimal payoffs (cf. [16, Ch. 2]). Notice, however, that minimization is in this case trivial as all payoffs always have the same value: strictly positive if $\sigma_i^{\mathcal{A}}$ is winning, 0 otherwise.

[14] Theorem 1 is the special case of Theorem 2 where the belief space is a singleton and $B = \langle 1 \rangle$.

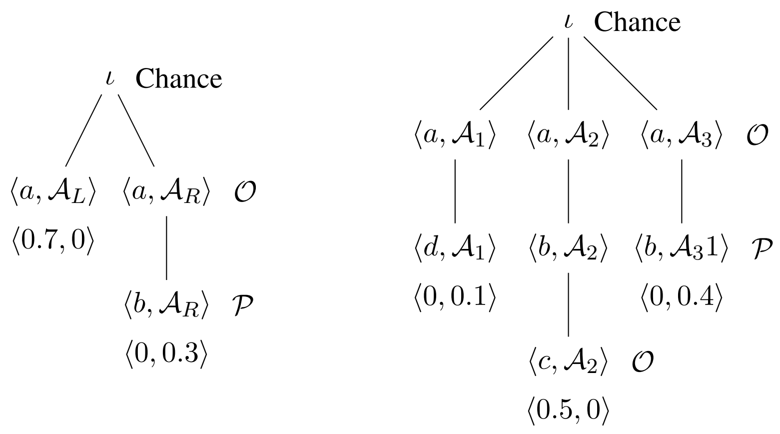

Fig. 5: The rendering of Example 3 (left) and Example 4 (right) as argument games with perfect information over one initial chance move: the two politicians expecting to be interviewed by a pool of known journalists; and a prosecutor and defense lawyer expecting a judge to be drawn for the trial from a pool of known judges.

5.2 Imperfect information on the initial chance move

Example 7 (Example 6 continued)**.** *Assume now the two politicians will be interviewed, as above, by a group of journalists representative of the population's distribution* $\langle 0.3, 0.7 \rangle$*. However, this time, they do not know the types of these journalists. How will the debate proceed?*

In the example proponent and opponent do not know the type of their listener. In other words, although they know what the probability is of each graph, they do not know for sure with respect to which graph their dialogue will be evaluated by the listener.

5.2.1 The game

Two intuitions will back the set up of our model: first, if players do not have information about which graph has been picked by Chance, then players will argue by attempting any attack that would be compatible with at least one of the graphs in the belief space; second, they will argue 'by best bet' will put forth arguments according to what is believed to be the most plausible graph.

From a modeling point of view it should then be possible that the game generates dialogues that are incompatible with some of the underlying graphs. This motivates us to introduce the following auxiliary notion. For a graph $\mathcal{A}$ and dialogue $\mathbf{a}$, let $\mathbf{a}_{\mathcal{A}}$ denote the largest subsequence of $\mathbf{a}$ which is legal with respect to $\mathcal{D}(\mathcal{A})$, i.e., which belongs to its possible dialogues. Clearly, it can be the case that

$\mathbf{a}_{\mathcal{A}} = \mathbf{a}$. Define now the following functions:

$$l_{\mathcal{P}}(\mathbf{a}_{\mathcal{A}}) = \begin{cases} 1 & \text{IF } \mathtt{t}(\mathbf{a}_{\mathcal{A}}) = \mathcal{O} \text{ AND } \ell(\mathbf{a}_{\mathcal{A}}) < \omega \\ 0 & \text{OTHERWISE} \end{cases}$$

$$l_{\mathcal{O}}(\mathbf{a}_{\mathcal{A}}) = \begin{cases} 0 & \text{IF } \mathtt{t}(\mathbf{a}_{\mathcal{A}}) = \mathcal{O} \text{ AND } \ell(\mathbf{a}_{\mathcal{A}}) < \omega \\ 1 & \text{OTHERWISE} \end{cases}$$

Intuitively, the function records when a player has had 'the last word' in the largest dialogue in $\mathbf{a}$ compatible with $\mathcal{A}$, keeping the asymmetry typical of the payoff function of the game for grounded (recall Table 1).

Example 8. *Recall Example 3 and Figure 3. Consider sequence ab. We have that $ab_{\mathcal{A}_R} = ab$ and $ab_{\mathcal{A}_L} = a$. As to the 'last word' function we have: $v_{\mathcal{P}}(ab_{\mathcal{A}_L}) = 0$ and $v_{\mathcal{O}}(ab_{\mathcal{A}_L}) = 1$; $v_{\mathcal{P}}(ab_{\mathcal{A}_R}) = 0$ and $v_{\mathcal{O}}(ab_{\mathcal{A}_R}) = 1$.*

Definition 10 (Imperfect information argument games)**.** *The imperfect information argument game is a function $\mathcal{D}^I(.)$ which for each belief B yields structure $\mathcal{D}^I(B) = \langle N^I, A^I, \mathtt{t}^I, \mathtt{m}^I, \mathtt{p}^I \rangle$ where:*

- *$N^I, A^I, \mathtt{t}^I$ are as in Definition 4.*
- *$\mathtt{m}^I$ is defined as:*

  $$\mathtt{m}^I(\mathbf{a}) = \left\{ a \in A \mid \exists \mathcal{A} \in \mathrm{A} : a \in \mathtt{m}^{\mathcal{A}}(h(\mathbf{a})) \right\}.$$

  where $\mathtt{m}^{\mathcal{A}}$ denotes the move function in the game for grounded on graph $\mathcal{A}$. I.e., the available moves at $\mathbf{a}$ are those arguments which may be moved in at least one of the graphs in A. As usual, a dialogue of length ω or such that $\mathtt{m}^I(\mathbf{a}) = \emptyset$ is called terminal. The set of terminal dialogues is denoted Z.
- *$\mathtt{p}^I : N \times Z \longrightarrow [0,1]$ is the payoff function defined as follows:*

  $$\mathtt{p}(i)(\mathbf{a}) = \sum_{\mathcal{A} \in \mathrm{A}} B(\mathcal{A}) \cdot l_i(\mathbf{a}_{\mathcal{A}}) \tag{2}$$

  I.e., the payoff for i in a terminal dialogue $\mathbf{a}$ is determined by how often player i expects—according to B—to have the last word in $\mathbf{a}$.

For $a \in A$, $\mathcal{D}^I(B)@a$ denotes the imperfect information argument game instantiated at a.

Here are two examples. The game on the right of Figure 6 is the game modeling Example 7. The game in Figure 7 models the imperfect information version of Example 4 where the plaintiff and defense layer do not know the identity of the judge that will be selected for the trial.

A few observations: the move function allows arguers to put forth any argument that attacks the last uttered one in some of the available attack graphs; the payoff function is designed to ignore, for each graph in the belief space, those parts of dialogues that are illegal with respect to the given graph; the game is still a zero-sum game, although payoffs range over $[0,1]$ instead of $\{1,0\}$ (recall Remark 1).

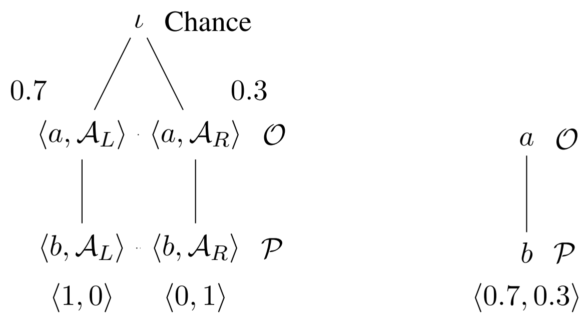

Fig. 6: The rendering of Example 7 by a canonical imperfect information game with information sets—the dotted lines—(left), and its representation as the game of Definition 10 (right). Notice that the payoffs of histories of information sets, i.e., the weighted sum of all the histories in the information set (e.g., $1 \cdot 0.7 + 0 \cdot 0.3$), coincide with the payoffs of our game computed by function $\mathtt{p}^P$: $\mathtt{p}^P_{\mathcal{P}}(ab) = 0.7$.

Remark 4 (Imperfect information). *Readers versed in game theory must have noticed that we have introduced the game of this section as a perfect information game but, technically, we have modeled it in Definition 10 as a perfect information game. In fact, we have not used any of the machinery typically employed in game theory to account for imperfect information (e.g., information sets [16, Ch. 11]). This modeling solution is possible because of the peculiar sort of uncertainty we are studying, which is due to one initial Chance move and which is the same for each player. This allows us to collapse information sets (i.e., all pairs $\langle a, \mathcal{A} \rangle$ for a given a) to a single arguments (a itself) and, consequently, histories consisting of information sets to sequences of arguments. Hence, the game form is essentially the same. Finally, our payoff function $\mathtt{p}^P$ calculates payoffs precisely as the weighted sum of the payoffs of the dialogues that would constitute a history of information set. Figure 6 gives a concrete example of this correspondence.*

5.2.2 Strategies

We define now strategies and their values.

Definition 11. *Let $\mathcal{D}^I(B)@a$ be an imperfect information game for belief B instantiated at a:*

- *Strategies are defined as in Definition 5. Like in the case of the game for grounded, a pair of strategies $\langle \sigma_{\mathcal{P}}, \sigma_{\mathcal{O}} \rangle$ describes a terminal dialogue. We denote the set of strategies available to player i, Σ_i.*

- *The value of strategy $\sigma_{\mathcal{P}}$ is given by:*

$$\min_{\sigma_{\mathcal{O}} \in \Sigma_{\mathcal{O}}} (\mathtt{p}^I(\langle \sigma_{\mathcal{P}}, \sigma_{\mathcal{O}} \rangle))$$

  I.e., the value of a strategy $\sigma_{\mathcal{P}}$ equals the minimum payoff that $\mathcal{P}$ can obtain via that strategy under all possible $\mathcal{O}$'s responses.[15] *The same definition applies for $\mathcal{O}$.*

- *The set of arguments for which i has a strategy with value* at most *α is denoted $Win_i^{\alpha}(\mathcal{D}^P(B))$.*

5.2.3 Adequacy failure

Ideally, we would now like to show that the imperfect information game is still adequate with respect to a probabilistic version of the grounded extension, along the line of Theorem 2: $\alpha = p_B(a) \iff a \in Win_{\mathcal{P}}^{\alpha}(\mathcal{D}^I(B))$. However, it turns out that this is not the case. A counterexample can be obtained by building the game for Example 4, and is given in Figure 7.

This failure is due to two concurrent factors. The first is, clearly, the imperfect information assumption over the moves by Chance. The second has to do with a structural constraints of the game for grounded, namely the fact that players can move only one argument at the time. This means that players must sometimes 'sacrifice' good arguments simply because they are less probable (e.g., $\mathcal{O}$ in Figure 7).

We conjecture that adequacy can be restored for the imperfect information game while retaining it in the other games, for instance by allowing players move sets of arguments (see [10]). The technical implications of such a more general set-up is left for future work.

6 Conclusions and future work

Starting from the argument game for the grounded extension, the paper has discussed a probabilistic model of uncertainty in argumentation which lead to a probabilistic version of the grounded extension, and to the specification of two different types of argument games modeling two different ways uncertainty influences the game for grounded. The paper has also discussed a specifically game-theoretic view on adequacy results for argument games and, after proving the adequacy of the game for grounded (Theorem 1), it has established the same result for the probabilistic grounded extension (Theorem 2), and has shown how and why adequacy is disrupted in imperfect information argument games.

Future work will aim at relaxing the assumption that proponent and opponent share one same belief. Dropping this assumption would lead, in the case of imperfect information, to variable-sum games. Furthermore, more refined techniques

[15] This is the idea behind the game-theoretic notion of maximinimization [16, Ch. 2] and is naturally applicable to all zero-sum games.

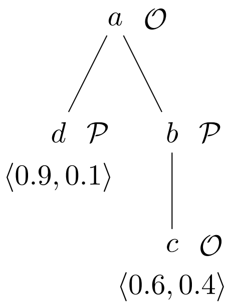

Fig. 7: Although the probability of a belonging to the grounded extension is 0.5, $\mathcal{P}$ has in a a strategy with value 0.6. This is due to the fact that $\mathcal{O}$ has to choose whether to play d or b. Choosing b he guarantees himself a better payoff (0.4 instead of 0.1), but still lower than his winning probability (under perfect information) of 0.5.

for epistemic modeling could be employed to support players reasoning about each others' beliefs.

References

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

[2] P. Baroni and M. Giacomin. Semantics of abstract argument systems. In I. Rahwan and G. R. Simari, editors, *Argumentation in Artifical Intelligence*. Springer, 2009.

[3] T. Bench-Capon, S. Doutre, and P. E. Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171:42–71, 2007.

[4] T. Bench-Capon and P. E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10):619–641, 2007.

[5] M. Caminada and G. Pigozzi. On judgment aggregation in abstract argumentation. *Journal of Autonomous Agents and Multi-Agent Systems*, 22(1):64–102, 2011.

[6] S. Coste-Marquis, C. Devred, S. Konjeczny, M.-C. Lagasquie-Schiex, and P. Marquis. On the merging of Dung's argumentation systems. *Artificial Intelligence*, 171:730–753, 2007.

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

[8] P. M. Dung. Logic programming as dialogue games. Technical report, Division of Computer Science, Asian Institute of Technology, 1994.

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

[10] P. M. Dung and P. M. Thang. A unified framework for representation and development of dialectical proof procedures in argumentation. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*, pages 746–751, 2009.

[11] J. Glazer and A. Rubinstein. Debates and decisions: On a rationale of argumentation rules. *Games and Economic Behavior*, 36(2):158–173, 2001.

[12] A. Hunter. Towards higher impact argumentation. In *Proceedings of the 19th American Conference on Artificial Intelligence (AAAI'04)*, pages 275–280. AAAI Press, 2004.

[13] H. Jacobovits and D. Vermeir. Dialectic semantics for argumentation frameworks. In *Proceedings of the 7th International Conference on Artificial Intelligence and Law (ICAIL'99)*, pages 53–62, 1999.

[14] P. Matt and F. Toni. A game-theoretic measure of argument strength for abstract argumentation. In *Proceedings of 11th European Conference on Logics in Artificial Intelligence (JELIA'08)*, volume 5293 of *LNAI*, pages 285–297, 2008.

[15] S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G. Simari, editors, *Argumentation in AI*, pages 105–132. Springer, 2009.

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

[17] H. Prakken. Modeling defeasibility in law: Logic or procedure. *Fundamenta Informaticae*, 20:1–20, 2001.

[18] H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:187–219, 2001.

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

[20] I. Rahwan, K. Larson, and F. Tohme. Characterisation of strategy-proofness for grounded argumentation semantics. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*, pages 251–256, 2009.

[21] R.Parikh. Social software. *Synthese*, 132(3):187–211, 2002.

[22] P. A. Tohmé, G. A. Bodanza, and G. R. Simari. Aggregation of attack relations: A social-choice theoretical analysis. In S. Hartmann and G. Kern-Isberner, editors, *Foundations of Information and Knowledge Systems (FolKS 2008)*, pages 8–23, 2008.

[23] G. Vreeswijk and H. Prakken. Credulous and sceptical argument games for preferred semantics. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'00)*, LNAI, pages 239–253. Springer, 2000.

[24] E. Zermelo. Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In *Proceedings of the 5th Congress Mathematicians*, pages 501–504. Cambridge University Press, 1913.