Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence

Audience-Based Uncertainty in Abstract Argument Games

Davide Grossi and Wiebe van der Hoek
Department of Computer Science
University of Liverpool
{d.grossi,wiebe.van-der-hoek}@liverpool.ac.uk

Abstract

The paper generalizes abstract argument games to cope with cases where proponent and opponent argue in front of an audience whose type is known only with uncertainty. The generalization, which makes use of basic tools from probability theory, is motivated by several examples and delivers a class of abstract argument games whose adequacy is proven robust against uncertainty.

1 Introduction

Abstract argumentation theory [Dung, 1995] 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 work in the area has been devoted to defining and studying such standards of proof—the so-called extensions—in terms of structural properties of attack graphs (see [Baroni and Giacomin, 2009]). Some research has then focused on operationalizations of those definitions via two-player games (see [Modgil and Caminada, 2009]).

Although being evident in real-life arguments, debates, and negotiations, the strategic dimension of argumentation has been object to only limited scrutiny in the literature on abstract argumentation games. In this paper we investigate the sort of strategic issues that arise in abstract argument games when the resulting argumentation—i.e., the arguments put forth during the game—is evaluated by an audience whose type the arguers know only with uncertainty.

Scientific context In recent years, techniques from economic theory have found their way to abstract argumentation but, interestingly enough, there has been a clear bias in favor of techniques coming from the field of social choice theory and mechanism design targeting the problem of 'aggregation' in abstract argumentation [Coste-Marquis *et al.*, 2007; Tohmé *et al.*, 2008; Rahwan *et al.*, 2009; Caminada and Pigozzi, 2011; Bonzon and Maudet, 2011].

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 or dialogue 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 belongs to a given extension. This property is called *adequacy*.

Within the tradition initiated by Dung, besides the abovementioned literature, we are only aware of: [Matt and Toni, 2008], which studies a strategic zero-sum game with randomization, using it to obtain quantitative refinements of some of Dung's semantics; and [Procaccia and Rosenschein, 2005], which studies games of negotiation based on Dung's formalism. A few game theorists have also taken up the challenge of modeling some features of real-life debates. In particular, [Glazer and Rubinstein, 2001] studies a game by which two speakers aim at persuading a third party—which is also the general scenario of the current paper—but it is framed at an even higher level of abstraction than the one enabled by Dung's framework. Moreover, none of these works addresses the adequacy issue with respect to a given Dung extension.

Recently, uncertainty has attracted some attention in abstract argumentation [Li *et al.*, 2011; Hunter, 2012], where Dung's theory is interfaced with probability theory. However, to the best of our knowledge, no work to date has studied the games that a probability-based representation of uncertainty in abstract argumentation is able to support.

Contribution of the paper The paper shows how the current theory of abstract argument games can be adapted to provide an analysis of games where players aim at persuading an audience witnessing the argument. By persuasion we mean here persuasion of a third party in the argumentation, not persuasion of one of the arguers by another arguer.[1]

The presence of an audience introduces a specific type of uncertainty in the debaters which influences their strategic choices in the argumentation. We are interested in studying whether argument games under that type of uncertainty may remain adequate with respect to given Dung-style extensions and, therefore, remain 'fair' for proponent and opponent. Concretely, this audience-induced uncertainty is modeled as uncertainty about the underlying Dung graphs upon which the argument games are played.

Structure of the paper Section 2 introduces some necessary background from abstract argumentation and Section 3

[1]The latter is the setup of another strand of research in (instantiated) argumentation: the theory of *persuasion dialogues* (cf. [Prakken, 2009]).

recapitulates some known results about argument games emphasizing their game-theoretic component. Section 4 introduces our probabilistic model of audiences and proves a simple probabilistic adequacy theorem. Section 5 introduces abstract persuasion games and proves their probabilistic adequacy. Related work is discussed throughout, and conclusions follow in Section 6.

2 Abstract argumentation

2.1 Attack graphs

We start by introducing the notion of attack graph, or argumentation framework, due to [Dung, 1995]:

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* $\mathfrak{A}(A)$. *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* $R_{\mathcal{A}}(a)$ *the set of arguments attacking* a: $\{b \in A \mid b \rightarrow a\}$.

The set A represents a set of atomic arguments. Relation $\rightarrow$, called the 'attack' relation, encodes the way arguments attack one another. In short, an attack graph is a high-level representation of the conflicts inherent within the information put forth by a set of arguments.

2.2 Solving attack graphs

By 'solving' an attack graph we mean selecting a subset of arguments that enjoy some characteristic structural property. Several of these properties—known as *extensions*—have been proposed in the literature (cf. [Baroni and Giacomin, 2009]). In this paper we focus on the so-called grounded extension. This choice is dictated by two reasons: first, the grounded extension has a compact mathematical formulation (in terms of fixpoint theory) which is functional to a speedy exposition of the technical parts of this paper; second, it has a well understood operationalization in terms of games. It must be stressed, however, that the methods developed here could be adapted to any extension definable through two-player zero-sum games (cf. [Vreeswijk and Prakken, 2000; Dung and Thang, 2007]).

The grounded extension is defined via a function, called *characteristic function*, which is defined for every attack graph $\mathcal{A} = \langle A, \rightarrow \rangle$ as: $\mathtt{d}_{\mathcal{A}}(X) = \{a \in A \mid \forall b \in A : \text{ IF } b \rightarrow a \text{ THEN } X \rightarrow b\}$.

The function outputs, for each set of arguments X, the set of arguments defended by X. A set of arguments X such that $X \subseteq \mathtt{d}_{\mathcal{A}}(X)$ is therefore able to defend itself from external attacks. The *grounded extension* of an attack graph $\mathcal{A}$ is defined as the least fixpoint of $\mathtt{d}_{\mathcal{A}}$, in symbols: $\mathsf{lfp}(\mathtt{d}_{\mathcal{A}})$. By the Knaster-Tarski theorem and the fact that the characteristic function is monotonic, the least fixpoint of the characteristic function exists for any attack graph and equals the intersection of all pre-fixpoints of the characteristic function, i.e., $\bigcap \{X \subseteq A \mid \mathtt{d}_{\mathcal{A}}(X) \subseteq X\}$.

Example 1. *In Figure 1 (left), writing* $\mathtt{d}$ *for* $\mathtt{d}_{\mathcal{A}}$, *we obtain* $\mathtt{d}(\emptyset) = \{c\}$, $\mathtt{d}(\{c\}) = \mathtt{d}^2(\emptyset) = \{a, c\}$ *and* $\mathtt{d}(\{a, c\}) = \mathtt{d}^3(\emptyset) = \{a, c\}$, *which is the grounded extension. On the right, we have:* $\mathtt{d}(\emptyset) = \emptyset$, *which is the grounded extension.*

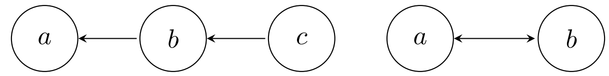

Figure 1: Two simple examples of attack graphs.

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

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

3 Argument games

The section recapitulates key definitions and results pertaining to an adequate game for the grounded extension. Our exposition, which differs from standard ones like [Modgil and Caminada, 2009], frames argument games as two-player, zero-sum, perfect-information extensive games with (possibly) infinite horizon.

3.1 Game for the grounded extension

This section introduces an argument game for the grounded extension (*game for grounded*, in short).

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

- $N = \{\mathbf{P}, \mathbf{O}\}$—*the set of players consists of proponent* $\mathbf{P}$ *and opponent* $\mathbf{O}$.
- A *is the set of arguments in* $\mathcal{A}$.
- $\mathtt{t} : A^{<\omega} \longrightarrow N$ *is the* turn function. *It is a (partial[2]) 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* $\mathtt{t}(\mathbf{a}) = \mathbf{O}$, *and if* $\ell(\mathbf{a}) = 2m + 1$ *then* $\mathtt{t}(\mathbf{a}) = \mathbf{P}$. *I.e., even positions are assigned to* $\mathbf{O}$ *and odd positions to* $\mathbf{P}$.
- $\mathtt{m} : A^{<\omega} \longrightarrow \wp(A)$ *is a (partial) function from dialogues to sets of arguments defined as:* $\mathtt{m}(\mathbf{a}) = R_{\mathcal{A}}(h(\mathbf{a}))$. *I.e., the available moves at* $\mathbf{a}$ *are the arguments attacking the last argument of* $\mathbf{a}$. *The set of all dialogues compatible*

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

with 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 in Table 1 associating 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, $\mathcal{D}(\mathcal{A})@a$).

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

3.2 Adequacy

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

Definition 3 (Strategies). *Let $\mathcal{D}(\mathcal{A}) = \langle N, A, \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$, $a \in A$ and $i \in N$. A strategy for i in the instantiated game $\mathcal{D}@a$ is a function: $\sigma_i : \{\mathbf{a} \in D - T \mid \mathbf{a}_0 = a \text{ AND } \mathtt{t}(\mathbf{a}) = i\} \longrightarrow A$ s.t. $\sigma_i(\mathbf{a}) \in \mathtt{m}(\mathbf{a})$. The set of terminal dialogues compatible with σ_i is defined as follows: $T_{\sigma_i} = \{\mathbf{a} \in T \mid \forall n \leq \ell(\mathbf{a}) \text{ IF } \mathtt{t}(\mathbf{a}|_n) = i \text{ THEN } \mathbf{a}_{n+1} = \sigma_i(\mathbf{a}|_n)\}$.*

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

Definition 4 (Winning strategies and arguments). *Let $\mathcal{D}(\mathcal{A}) = \langle N, S, \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$, $a \in A$ and $i \in N$. A strategy σ is* winning for i *in $\mathcal{D}(\mathcal{A})@a$ 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 $\mathcal{D}(\mathcal{A})@a$. The set of winning positions of $\mathcal{D}$ for i is denoted $Win_i(\mathcal{D}(\mathcal{A}))$.*

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 $\mathbf{P}$ or $\mathbf{O}$ possesses a winning strategy, and hence that each argument in an attack graph is either a winning position for $\mathbf{P}$ or a winning position for $\mathbf{O}$. See Figure 2 for an illustration.

We can now state the adequacy result for the game:

Theorem 1 (Adequacy of $\mathcal{D}$ [Dung, 1994]). *Let tuple $\mathcal{D}(\mathcal{A}) = \langle N, S, \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$ be the dialogue game for grounded on graph $\mathcal{A}$ and $a \in A$: $a \in \mathsf{lfp}(\mathtt{d}_{\mathcal{A}}) \iff a \in Win_{\mathbf{P}}(\mathcal{D}(\mathcal{A}))$.*

In the rest of the paper we will sometimes refer to $\mathcal{D}(.)$ as the *basic game* for grounded.

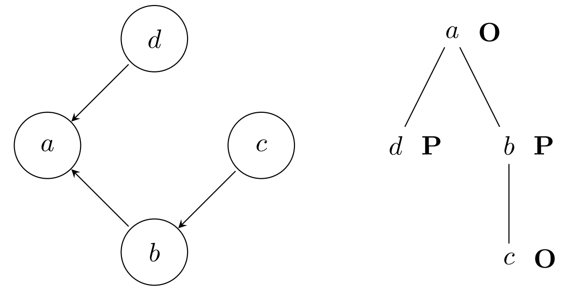

Figure 2: An attack graph (left) and its dialogue game for grounded (right). Positions are labeled by the player whose turn it is to play. $\mathbf{P}$ wins the terminal dialogue abc but loses the terminal dialogue ad. $\mathbf{O}$ has a winning strategy.

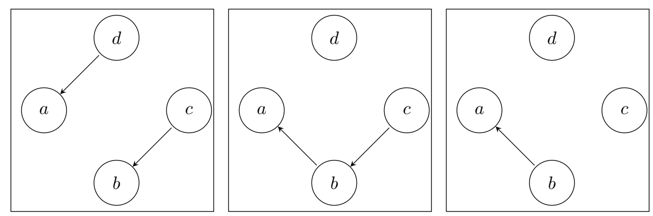

Figure 3: The belief space of Example 3. We will refer to the three graphs as $\mathcal{A}_1, \mathcal{A}_2$ and $\mathcal{A}_3$.

4 Audiences and uncertainty

In this section we propose a probabilistic model of the sort of uncertainty that arises when arguers exchange arguments in front of a listener whose attitude towards their arguments they are uncertain about, or in front of an audience composed by different types of listeners.[3]

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 $\mathbf{A}$ (the belief space).[4] 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 we use the term *belief* in a technical sense as the representation of what arguers think the probabilities are of each graph to be extracted by a lottery. This non-committal reading applies to many situations we wish to model. Here are two examples:

Example 2. *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*

[3] Audiences and persuasion have been recent object of study, albeit not from a game theoretic point of view, in [Hunter, 2004; Bench-Capon *et al.*, 2007; Oren *et al.*, 2012].

[4] Beliefs can be related to *probabilistic argumentation frameworks* [Li *et al.*, 2011; Hunter, 2012; 2013], where each argument is endowed with a value in the $[0, 1]$ interval.

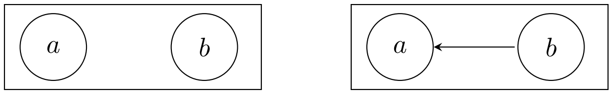

Figure 4: The belief space of Example 2. $\mathcal{A}_L$ will denote the left graph and $\mathcal{A}_R$ the right one.

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 4). The politicians' belief are modeled by distribution* $B = \langle 0.7, 0.3 \rangle$.

Example 3. *Assume again the arguments* a, b, c *and* d *of Figure 2. A prosecutor and a defense lawyer are arguing about argument* a*. They are uncertain about how the judge is going to interpret the evidence at hand. They believe that it is* 10% *probable that the judge would not recognize* b *as legitimate evidence against* a*, while accepting the other 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 3 and the belief is* $B = \langle 0.1, 0.5, 0.4 \rangle$.

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 $\mathbf{A}$: $g(\mathcal{A})(a) = \begin{cases} 1 & \text{IF } a \in \mathsf{lfp}(\mathsf{d}_{\mathcal{A}}) \\ 0 & \text{OTHERWISE} \end{cases}$.
We can formulate the following definition:

Definition 5 (Probability of being grounded). *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 computed by the following formula:*

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

If $p_B(a) = \alpha$*, we say that* a *is grounded with probability* α.

Obviously, $1 - p_B(a)$ denotes 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 4. *Recall Example 2. We have quite simply that* $p_B(a) = 0.7$ *and* $p_B(b) = 1$*. For Example 3, we have* $p_B(a) = 0.5$*,* $p_B(b) = 0.4$ *and* $p_B(c) = p_B(d) = 1$.

Probabilistic adequacy

Suppose now to randomly select, given a belief $\Delta(\mathbf{A})$, an attack graph from $\mathbf{A}$. There is a precise sense in which this random selection (or lottery) is adequate with respect to the probabilistic generalization of the grounded extension given in Definition 5. Let $\mathbf{A} \subseteq \mathfrak{A}(A)$ be a belief space and let $Win_i^{\alpha}(\mathcal{D}(\mathbf{A}))$ denote the set of arguments in A from which i has a winning strategy with probability α.

Theorem 2 (Probabilistic adequacy). *Let* $B = \Delta(\mathbf{A})$*,* a *be an argument and* $0 \leq \alpha \leq 1$*:* $p_B(a) = \alpha \iff a \in Win_{\mathbf{P}}^{\alpha}(\mathcal{D}(\mathbf{A}))$.

Proof. Let $\mathbf{A}_a = \{\mathcal{A} \in \mathbf{A} \mid a \in Win_{\mathbf{P}}(\mathcal{D}(\mathcal{A}))\}$. The result follows by this series of equivalences:

$$\begin{aligned} p_B(a) = \alpha &\iff \sum_{\mathcal{A} \in \mathbf{A}} B(\mathcal{A}) \cdot g(\mathcal{A})(a) = \alpha \\ &\iff \sum_{\mathcal{A} \in \mathbf{A}_a} B(\mathcal{A}) = \alpha \\ &\iff a \in Win_{\mathbf{P}}^{\alpha}(\mathcal{D}(\mathbf{A})) \end{aligned}$$

The first equivalence holds by Definition 5, the second by Theorem 1, the third by the definition of $Win_{\mathbf{P}}^{\alpha}(\mathcal{D}(\mathbf{A}))$. □

Intuitively, given a belief B, an argument belongs to the grounded extension with probability α if and only if α is the probability by which $\mathbf{P}$ can guarantee to win the game for grounded on a graph selected at random given the probability distribution in B. A way to picture this is that $\mathbf{P}$ is faced with a game in two stages: first, a third player Chance selects an attack graph according to the probability distribution of B; second, $\mathbf{P}$ plays the game for grounded against $\mathbf{O}$ on the selected graph.

Example 5 (Example 2 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 4 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.*

5 Abstract games of persuasion

The section presents an abstract argument game—called *persuasion game*—probabilistically adequate with respect to the grounded extension (Definition 5). This is achieved in two steps. First, a naive extension of the argument game for grounded (Definition 2) is presented to cope with beliefs, which is shown not to be adequate. Second, based on the insights gained, a new game is presented and proven adequate.

5.1 Persuasion based on the game for grounded

Example 6 (Example 5 continued). *Assume now the two politicians will be interviewed, as above, by one journalist from a pool of journalists with type distribution* $\langle 0.7, 0.3 \rangle$*. However, this time, the journalists in the pool are all unknown to them, so they will be unable to recognize the type of the selected journalist. The only information they have is the type distribution. How will the debate proceed?*

In the example $\mathbf{P}$ and $\mathbf{O}$ do not know the type of their listener (the selected journalist). In other words, although they know what the probability is of each graph, they do not know for sure with respect to which graph precisely their dialogue will be evaluated by the listener. Bringing in, as we did in Section 4.1, a third player Chance, this situation can be thought of as $\mathbf{P}$ and $\mathbf{O}$ having to argue without knowing which graph Chance has selected.

The game

Two intuitions 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' and will put forth arguments according to what are believed to be the most probable graphs. It is then 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 (prefix) 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:

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

The function $l_{\mathbf{O}}(\mathbf{a}_{\mathcal{A}})$ for $\mathbf{O}$ is defined accordingly by inverting the cases for the assignment of 1 and 0. Intuitively, the functions record when a player has had 'the last word' in the largest dialogue in $\mathbf{a}$ compatible with $\mathcal{A}$.

Example 7. *Recall Example 2 and Figure 4. 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: $l_{\mathbf{P}}(ab_{\mathcal{A}_L}) = 1$ and $l_{\mathbf{O}}(ab_{\mathcal{A}_L}) = 0$; $l_{\mathbf{P}}(ab_{\mathcal{A}_R}) = 0$ and $l_{\mathbf{O}}(ab_{\mathcal{A}_R}) = 1$.*

Definition 6 (Persuasion game). *The persuasion game is a function $\mathcal{P}$, which for each belief B over $\mathbf{A} \subseteq \mathfrak{A}(A)$ yields a structure $\mathcal{P}(B) = \langle N, A, \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$ where:*

- *$N, A, \mathtt{t}$ are as in Definition 2.*
- *$\mathtt{m}$ is defined as: $\mathtt{m}(\mathbf{a}) = \{a \in A \mid \exists \mathcal{A} \in \mathbf{A} : a \in \mathtt{m}^{\mathcal{A}}(h(\mathbf{a}))\}$ 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 $\mathtt{m}(\mathbf{a}) = \emptyset$ is called terminal. The set of terminal dialogues is denoted T.*
- *$\mathtt{p} : N \times T \longrightarrow [0, 1]$ is the payoff, so defined:*

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

  I.e., the payoff for i in a terminal dialogue consists in how often i expects to have the last word in $\mathbf{a}$.

For $a \in A$, $\mathcal{P}(B)@a$ is the persuasion game starting at a.

The game on the left side of Figure 5 models the persuasion game for Example 3 where the plaintiff and defense lawyer do not know the identity of the judge 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 $\{1, 0\}$.

Strategies and adequacy failure

We define strategies and their values, i.e., the best payoff players can secure themselves by playing a given strategy.

Definition 7. *Let $\mathcal{P}(B)@a$ be a persuasion game for belief B instantiated at a:*

- *Strategies are defined as in Definition 3. Like in the case of the game for grounded, a pair of strategies $(\sigma_{\mathbf{P}}, \sigma_{\mathbf{O}})$ describes a terminal dialogue. We denote the set of strategies available to player i, Σ_i.*
- *The value of strategy $\sigma_{\mathbf{P}}$ is $\min_{\sigma_{\mathbf{O}} \in \Sigma_{\mathbf{O}}}(\mathtt{p}(\mathbf{P})(\sigma_{\mathbf{P}}, \sigma_{\mathbf{O}}))$. I.e., the value of a strategy $\sigma_{\mathbf{P}}$ equals the minimum payoff that $\mathbf{P}$ can obtain via that strategy under all possible $\mathbf{O}$'s responses. The same applies for $\mathbf{O}$.*
- *The set of arguments for which $\mathbf{P}$'s best strategy has value α, i.e., $\max_{\sigma_{\mathbf{P}} \in \Sigma_{\mathbf{P}}} \min_{\sigma_{\mathbf{O}} \in \Sigma_{\mathbf{O}}}(\mathtt{p}(\mathbf{P})(\sigma_{\mathbf{P}}, \sigma_{\mathbf{O}})) = \alpha$ is denoted $Win^{\alpha}_{\mathbf{P}}(\mathcal{P}(B))$. The same applies for $\mathbf{O}$.*

We now address the following question: Is it possible to establish adequacy results for the persuasion game along the lines of Theorem 1 and Theorem 2? Ideally, one would like the persuasion game to be such that the payoff that Proponent can guarantee for herself is exactly the probability that the argument at which the game is instantiated belongs to the grounded extension: $p_B(a) = \alpha \iff a \in Win^{\alpha}_{\mathbf{P}}(\mathcal{P}(B))$.

However, it turns out that this is not the case for the game defined in Definition 6. A counterexample can be obtained by building the game tree for Example 3 (left side of Figure 5). There, $\mathbf{P}$ has a best strategy that guarantees her a payoff of at least 0.6 (cf. last bullet in Definition 7) while the probability for a to belong to the grounded extension is 0.5.

This discrepancy between the values of best strategies from an argument a and the probability of a belonging to the grounded extension is due to two factors. The first is, clearly, the uncertainty about the underlying graph. The second has to do with a structural constraint of the game for grounded, namely the fact that players can use only one argument at the time. This means that players must sometimes 'sacrifice' good arguments just because they enable attacks in less probable graphs. In the above example, $\mathbf{O}$ 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 probability (0.5) of winning a game for grounded.

5.2 An adequate persuasion game

We now show how adequacy can be obtained by appropriately modifying the move function. The new game will be called *persuasion game for* grounded.

A persuasion game for the grounded extension

The crux of the new definition consists in making it possible for $\mathbf{P}$ and $\mathbf{O}$ to play many arguments at the same time—a feature used, with different purposes, in [Dung and Thang, 2009]—as if they were playing many basic games for grounded 'in parallel' on several graphs. As no confusion will arise, we maintain notation $\mathcal{P}$ to denote the new game.

Definition 8 (Persuasion game *for* grounded). *The persuasion game for grounded is a function $\mathcal{P}$, which for each belief B over $\mathbf{A} \subseteq \mathfrak{A}(A)$ yields a structure $\mathcal{P}(B) = \langle N, \wp(A), \mathtt{t}, \mathtt{m}, \mathtt{p}\rangle$ where N and $\mathtt{t}$ are as in Definition 6 and:*

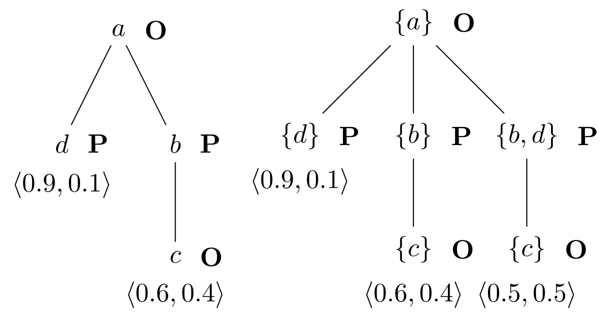

Figure 5: Left: game tree of the persuasion game for the graphs (belief space) in Figure 3. Right: game tree of the persuasion game *for* grounded for the same graphs.

- *$\wp(A)$ is the set of sets of arguments from A. The set of sequences of sets of arguments is S and the set of finite sequences of sets of arguments $S^{<\omega}$. Set S is therefore the set of all possible* dialogues *for this game. Elements of S are denoted $\mathbf{s}$. The n^{th} element of $\mathbf{s}$ is $\mathbf{s}_n$ for n positive. In addition, S is assumed to be s.t. $\forall \mathbf{s} \in S$, $\mathbf{s}_1$ is a singleton, i.e., dialogues start with only one argument. Denote by $d(\mathbf{s})$ the set $\{a \mid a \in \mathbf{s}_i \text{ FOR } 1 \leq i \leq \ell(\mathbf{s})\}$, i.e., the set of arguments in some element of $\mathbf{s}$. Finally, for $\mathcal{A} = \langle A, \rightarrow\rangle \in \mathbf{A}$ let $\mathcal{A}_\mathbf{s} = \langle A \cap d(\mathbf{s}), \rightarrow \cap (A \cap d(\mathbf{s}))^2\rangle$, i.e., the restriction of $\mathcal{A}$ to the set $d(\mathbf{s})$.*
- $\mathtt{m} : S^{<\omega} \longrightarrow \wp(\wp(A))$ *is defined as:* $\mathtt{m}(\mathbf{s}) = \{X \mid \forall a \in X, \exists \mathcal{A} \in \mathbf{A}, \exists \mathbf{a} \in d(\mathbf{s}) : a \in \mathtt{m}^{\mathcal{A}}(h(\mathbf{a}))\}$. *where $\mathtt{m}^{\mathcal{A}}$ denotes the move function in the game for grounded on graph $\mathcal{A}$. I.e., the moves at $\mathbf{s}$ are* sets of *arguments that can be played in some dialogue in some graph in $\mathbf{A}$. The set of terminal dialogues is denoted T.*
- $\mathtt{p} : N \times T \longrightarrow [0, 1]$*, the payoff function, is so defined:*
$$\mathtt{p}(i)(\mathbf{s}) = \sum_{\mathcal{A} \in \mathrm{A}} B(\mathcal{A}) \cdot \rho_i^{\mathbf{s}}(\mathcal{A})$$
with $\rho_i^{\mathbf{s}}(\mathcal{A}) = \begin{cases} g(\mathcal{A}_\mathbf{s})(\mathbf{s}_1) & \text{IF } i = \mathbf{P} \\ 1 - g(\mathcal{A}_\mathbf{s})(\mathbf{s}_1) & \text{OTHERWISE} \end{cases}$.[5]
I.e., the payoff for i is determined by how often i will be able to win the standard game for grounded (Definition 2) on the graphs containing the arguments put forth in $\mathbf{s}$, starting at the initial argument $\mathbf{s}_1$. Intuitively, any time a player can force a win in the basic game on some graph defined by $\mathbf{s}$, that graph gives i a fraction of its final payoff equal to the probability of that graph.

Instantiations of this game are denoted by $\mathcal{P}(B)@\{a\}$.

The definitions of strategy, value of a strategy winning position are straightforward variants of Definition 7.

Adequacy
We are ready to state the main result of this section:

Theorem 3 (Probabilistic adequacy of $\mathcal{P}$). *Let $B = \Delta(\mathbf{A})$, a be an argument, $\mathcal{P}(B)$ be a persuasion game for grounded and $0 \leq \alpha \leq 1$: $p_B(a) = \alpha \Longleftrightarrow a \in Win^{\alpha}_{\mathbf{P}}(\mathcal{P}(B))$.*

[5]By $g(.)(.)$ (Section 4.1) and Theorem 1, ρ states whether i can win the basic game for grounded on $\mathcal{A}_\mathbf{s}$ instantiated at argument $\mathbf{s}_1$.

Sketch of proof. To simplify notation, let us refer to $Win^{\alpha}_{\mathbf{P}}(\mathcal{P}(B))$ by $Win^{\alpha}_{\mathbf{P}}$ and let $\max_{\sigma_i}$ stand for $\max_{\sigma_i \in \Sigma_i}$. Strategies apply to the instantiated persuasion game $\mathcal{P}(B)@\{a\}$. The claim is proven by these equivalences:

$$\begin{aligned} a \in Win^{\alpha}_{\mathbf{P}} &\Longleftrightarrow \max_{\sigma_\mathbf{P}} \min_{\sigma_\mathbf{O}} (\mathtt{p}(\mathbf{P})(\sigma_\mathbf{P}, \sigma_\mathbf{O})) = \alpha \\ &\Longleftrightarrow \max_{\sigma_\mathbf{P}} \min_{\sigma_\mathbf{O}} \left(\sum_{\mathcal{A} \in \mathrm{A}} B(\mathcal{A}) \cdot \rho_{\mathbf{P}}^{(\sigma_\mathbf{P}, \sigma_\mathbf{O})}(\mathcal{A}) \right) = \alpha \\ &\Longleftrightarrow \sum_{\mathcal{A} \in \mathbf{A}_a} B(\mathcal{A}) = \alpha \\ &\Longleftrightarrow p_B(a) = \alpha \end{aligned}$$

The key equivalence is the third one. Assume the best strategy for $\mathbf{P}$ has value α. This means that no matter what the strategy of $\mathbf{O}$ is, she will always be able to force the game on a sequence $\mathbf{s}$ such that the combined probability of all the frames $\mathcal{A}$ for which $\mathcal{A}_\mathbf{s}$ is such that $\mathbf{P}$ can force a win there, is equal to α. Denoting with $\mathbf{A}_a$ the set of such graphs, we therefore obtain that $\alpha = \sum_{\mathcal{A} \in \mathbf{A}_a} B(\mathcal{A}) = p_B(a)$. ☐

Intuitively, the upshot of the theorem is that the game of Definition 8 provides a sound and complete protocol for the persuasion of an audience whose constitution is known to the arguers as a probability distribution. In other words, by playing this game, Proponent (resp. Opponent) will be able to persuade *precisely* that much of the audience that holds graphs where the argument at issue belongs (resp. does not belong) to the grounded extension; or she (resp. he) will be able to persuade the listener precisely as often as the listener's type is such that it holds a graph where the argument is (resp. is not) grounded.

We have seen this was not the case for the game on the left of Figure 5. The right of Figure 5 depicts, for comparison, the persuasion game for grounded played on the same graphs (Figure 2). There, $\mathbf{O}$ has a best strategy guaranteeing him a payoff he was not able to obtain in the game on the left: 0.5.

Finally, observe that from Theorem 3 if follows, for $\mathbf{A} = \{\mathcal{A}\}$: $p_B(a) = 1 \Longleftrightarrow a \in Win^{1}_{\mathbf{P}}(\mathcal{P}(\{\mathcal{A}\}))$. That is, the persuasion game for grounded is adequate w.r.t. the grounded extension also in the non-probabilistic case. Like the basic game of Definition 2, the persuasion game is an adequate game-theoretic operationalization of the grounded extension. Unlike it, however, it is robust for probabilistic uncertainty.

6 Conclusions and future work

The paper has addressed the topic of the persuasion of an audience from a Dung-style abstract perspective. It has modeled audiences as probability distributions over sets of attack graphs thereby obtaining a probabilistic generalization of the grounded extension. It has then defined abstract persuasion games as extensions of standard argument games played assuming such probability distributions. The paper has finally shown how even for these games it is possible to obtain adequacy results in the tradition of argument games.

Future work will aim at relaxing two assumptions: that proponent and opponent share one and the same belief; that beliefs range on graphs built on the same set of arguments.

References

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

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

[Bonzon and Maudet, 2011] E. Bonzon and N. Maudet. On the outcomes of multiparty persuasion. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*, pages 47–54. IFAAMAS, 2011.

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

[Coste-Marquis *et al.*, 2007] 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.

[Dung and Thang, 2007] P. M. Dung and P. M. Thang. A sound and complete dialectical proof procedure for sceptical preferred argumentation. In *Proceedings of LPNMR Workshop on Argumentation and Non-Monotonic Reasoning (ArgNMR07)*, 2007.

[Dung and Thang, 2009] 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.

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

[Dung, 1995] 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.

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

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

[Hunter, 2012] A. Hunter. Some foundations for probabilistic abstract argumentation. In *Proceedings of COMMA'12*. IOS Press, 2012.

[Hunter, 2013] A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54:47–81, 2013.

[Li *et al.*, 2011] H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In *Proceedings of the first international conference on theory and applications of formal argumentation*, pages 1–16, 2011.

[Matt and Toni, 2008] 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.

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

[Oren *et al.*, 2012] N. Oren, K. Atkinson, and H. Li. Group persuasion through uncertain audience modeling. In *Proceedings of COMMA'12*. IOS Press, 2012.

[Prakken, 2009] H. Prakken. Models of persuasion dialogue. In I. Rahwan and G. Simari, editors, *Argumentation in Artificial Intelligence*, chapter 14. Springer, 2009.

[Procaccia and Rosenschein, 2005] A. Procaccia and J. Rosenschein. Extensive-from argumentation games. In *Proceedings of EUMAS 2005*, pages 312–322, 2005.

[Rahwan *et al.*, 2009] 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.

[Tohmé *et al.*, 2008] 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.

[Vreeswijk and Prakken, 2000] 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.