Heuristics in Argumentation: A Game-Theoretical Investigation.[1]

Régis RIVERET [a], Henry PRAKKEN [b], Antonino ROTOLO [a], Giovanni SARTOR [a,c]

[a] *CIRSFID, University of Bologna, Italy*

[b] *Department of Information and Computing Sciences, Utrecht University and Faculty of Law, University of Groningen, The Netherlands*

[c] *European University Institute, Law Department, Florence, Italy*

Abstract. This paper provides a game-theoretical investigation on how to determine optimal strategies in dialogue games for argumentation. To make our ideas as widely applicable as possible, we adopt an abstract dialectical setting and model dialogues as extensive games with perfect information where optimal strategies are determined by preferences over outcomes of the disputes. In turn, preferences are specified in terms of expected utility combining the probability of success of arguments with the costs and benefits associated to arguments.

Keywords. Argumentation, Game Theory, Expected Utility.

1. Introduction

Over the years many dialogue games for argumentation have been proposed to shed light on questions such as which conclusions are (defeasibly) justified, or how procedures for debate and conflict resolution should be structured to arrive at a fair and just outcome.

An issue which has not yet received much attention is the common sense observation that the outcome of a debate does not solely depend on the premises of a case, but also on the strategies that parties in a dispute actually adopt. The problem studied in this paper is how to determine optimal strategies in dialogue games for argumentation. In particular, we will focus on 'adjudication' debates. In many debates there are just two participants, who aim to persuade each other to adopt a certain opinion. On the contrary, in adjudication debates, a neutral third party (for example, a judge, a jury or an audience) is involved, who at the end of a debate must decide whether to accept the statements that the opposing parties have made during the debate. In such debates the opposing parties must make estimates about how likely it is that the premises of their arguments will be accepted by the third party, i.e., by the adjudicator. Moreover, we want to take into account that the opposing parties may have non-trivial preferences over the outcome of a debate, so that optimal strategies are determined by two factors: the probability of success of their arguments and the costs/benefits of such arguments.

[1] Supported by the EU projects ONE-LEX (Marie Curie Chair), ESTRELLA (IST-2004-027665), and ALIS (IST-2004-027968).

To make our ideas as widely applicable as possible, we formulate our ideas in the abstract dialectical setting of [6, 7]. This allows us to make as few assumptions as possible on the underlying logic and structure of arguments.

A specification of preferences will be provided in terms of expected utility, combining the probability (computed using [8]'s techniques) that an argument, being successful, brings about certain benefits, and the costs of that argument. More precisely, a probability distribution is assumed with respect to the adjudicator's acceptance of the parties' statements. This distribution determines the probability of the arguments' success, which is to be established on the basis also of the probabilities of success of its counterarguments. The probability of success of the argument is then used in combination with the benefits brought about by the argument (if successful) and of its costs in order to predict the expected utility of such an argument. For instance, the benefits brought about by an argument's success could be the compensation awarded to the successful party. The costs could include the sacrifice of disclosing certain information that the player would have preferred to keep secret, or the expenses required for obtaining the evidence needed for one of its premises (e.g. carrying out expensive laboratory tests). If each argument has an expected utility then each strategy may have a different utility. Game theory allows us to determine the optimal strategies for arguers, that is, the strategies related to the preferred expected utility.

Let us turn to an example (henceforth, "the flat example") to illustrate our approach. The example is a legal one, since legal disputes are typical examples of adjudication debates.

The proponent *Pro*, John Prator, is the new owner of a flat in Rome. The previous owner sold the flat to John Prator, for the symbolic amount of 10 euros, through a notary deed. The previous owner had signed with Rex Roll, the opponent *Opp*, a rental agreement for the flat, which should apply for two more years. John has an interest in kicking out Rex, since he received an offer from another party, who intends to buy the flat paying 300000 euros upon the condition that Rex leaves the flat. Rex Roll needs to stay two more years in Rome and refuses to leave the flat, as the rental fee was 500 euros per month, which is a convenient price (other flats in Rome have a rental fee of at least 600 euros per month). Hence, John sues Rex and asks the judge to impose to Rex to leave the flat. We assume that legislation states that any previously signed rental agreement still applies to the new owner when (1) a flat is donated, or (2) flat value is less than 150000 euros. John's main argument A runs as follows: we do not have a donation since the flat was paid, thus the rental agreement is no longer valid and so Rex Roll has no right to stay. The opponent may present argument C that paying a symbolic amount of money indeed configures a case of donation, and John may reply with argument E that it is not the case because the property transfer was a sale formalized by a notary deed. Alternatively, Rex may present the argument B that the market value of the flat is of 120000 euros and so the rental agreement is valid, whereas John may reply with D saying that he will pay within 10 days 210000 euros to the previous owner, thus amending the transfer deed in order that it indisputably be a sale concerning a good of a value greater than 150000 euros. Table 1 provides the probability, costs and benefits for each argument[2] So the problem is the following. According to the analysis of the case, which strategy to adopt?

[2] Note that for some arguments, which e.g. counterattack opponent's attacks, we assume that there is no specific benefits or costs: what they produce is just what is assigned to the main argument A. We also assume that, if opponent jointly plays B and C, this results in the same benefits disjointly produced by B and C.

This paper is organised as follows. In Section 2 we briefly recall the main notions of [6, 7]'s argument games on which our approach is based. Section 3 provides an interpretation in game theory for such argument games and the definition of optimal strategies. In Section 4, we investigate a specification of the expected utility of a strategy by combining the probability of success of arguments with their associated costs and benefits. In Section 5, related works are briefly discussed. In Section 6, the approach is recapitulated and future investigations are suggested.

2. The dialectical setting

The dialectical setting assumed in the paper follows the format of [6, 7]. To specify our dialectical framework we need to provide three sets of assumptions, concerning the logic, the game protocol and the argument games. With regard to the logic we assume the following:

1. Arguments have a finite nonempty set of premises and one conclusion.
2. There is a binary relation of defeat between arguments.

The logical assumptions are complemented with the following requirements concerning the game protocol:

1. An *argument game* is played by two players *Pro* and Opp.
2. Informally, a *move* in an argument game is a withdrawal or is an argument that defeats an argument previously moved by the other party (except the first move). Formally, a move is a tuple (pl, id, a, t) where *pl* is the player of the move, *id* is the move identifier (a natural number), a is the argument moved, and t is the identifier of the move's target. The first and withdrawal moves have a 'dummy' target, to reflect that they do not reply to another move. A withdrawal will be denoted by $\emptyset$. Below we will often simply speak of a move as an argument or a withdrawal, leaving the other three elements obvious from the context.
3. Player *Pro* does not repeat moves.
4. Each *turn* of an argument game consists of a withdrawal or a sequence of at most m arguments such that $m \geq 1$ (m is determined by a specific protocol). The first turn consists of a single argument or a withdrawal (i.e. no debate takes place).
5. The turn shifts after a player has made m moves in a row or earlier if the player to move explicitly indicates that she has ended her turn (which we will leave implicit below).
6. Each move other than the first one defeats its target.
7. If a move is *legal* then it satisfies all preceding conditions. A withdrawal move is always legal. Specific protocols can add further conditions and then turn the 'if' into 'if and only if'.

With regard to argument games we make the following assumptions:

1. A game *terminates* if a player withdraws. If the set of arguments is finite then each game terminates, since the proponent may not repeat arguments.
2. Each game induces a *reply tree*, which consists of the argument moves as nodes and their target relations as links. Note that target (or 'reply') links are, unlike defeat relations from the logic, always unidirectional. Suppose A defeats C, and

A and *B* defeat each other, such that the defeat graph is $C \leftarrow A \leftrightarrow B$: this may induce a game tree $C \leftarrow A \leftarrow B \leftarrow A$.

3. Reply trees can be labeled as follows: a node is *in* iff all its children are *out*; and a node is *out* iff it has a child that is *in*. (Informally, the leaves of the tree are trivially *in* and then we can work our way upwards through the tree to determine the status of all other nodes.) As proven by [7], this is formally a special case of status assignments to defeat graphs in which each defeat graph has a unique and complete status assignment. If no player (having the possibility to reply with an argument) withdraws, the set of arguments that are *in* corresponds to the unique argument extension in all of the semantics of [3].
4. An argument move *M* in a reply tree *T favours Pro* if *M* is *in*; otherwise *M* favours *Opp*.
5. A game is *won* by a player if at termination the initial move favours the player.

Given all these assumptions our aim is to define the optimal strategies for *Pro* and *Opp* in argument games. As a matter of fact, we will study this issue for a slightly different kind of game, which is induced by an argument game as defined above. We are in fact interested in optimal *turn* selection (instead of optimal argument selection), so the game to which we will apply the game-theoretical techniques is in fact an argument game consisting of turns, that is, sequences of arguments. So, we work with the following structures:

1. A defeat graph in which the nodes are arguments and the links are defeat relations; which is a declarative representation of a set of available arguments with their defeat relations. The graph is said to be declarative because it does not display in any way whether and when the arguments were stated in a dialogue;
2. A reply tree of a single-move argument game in which the nodes are arguments and the links are reply links;
3. A multi-move argument game which is a sequence of turns by two players *Pro* and *Opp*. Each turn consists of zero or more arguments;
4. A game tree of all possible turn games in which the nodes are turns and the links express their temporal order in a game.

A single terminated argument game based on the defeat graph of the arguments of our example, and its associated reply graph are provided in figure 1.

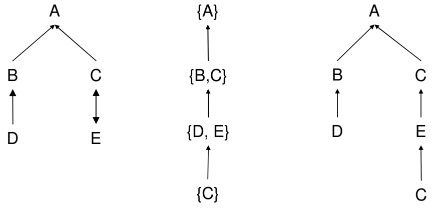

Figure 1. In the middle, a single terminated argument game based on the defeat graph on the left, and its reply graph on the right.

3. Game-theoretical model

Game theory deals with the heuristic layer of argumentation dialogue. For an observer, it helps to understand the moves of arguers in a dialogue. For an arguer, it helps to make the right moves by taking into account other arguers' behavior. In order to make the paper self-contained, we provide in the following the relevant game-theoretical notions from [5] slightly adapted to better fit our dialectical setting.

Most basic games presented in the literature are so-called *strategic games* that model situations in which all players' decisions are made simultaneously and each player chooses her plan of action once and for all. Modeling an argumentation dialogue as a strategic game is unsatisfactory because an arguer can plan moves not only at the beginning but also whenever she has to move. In other words, the model of strategic games does not allow arguers to reconsider which arguments to advance after some moves of the other parties. For this reason, we model dialogues as so-called *extensive games* to provide an explicit account of the sequential structure of argumentation. We also assume that arguers are *perfectly* informed about the arguments previously advanced by all other players. Bear in mind that games of perfect information in fact denote cases where no moves are simultaneous. Furthermore, we assume that the set of all arguments and their defeat relations (i.e. the defeat graph) is given in advance, is finite, stays fixed during a game and is known by both players during the game: in game-theoretical terms, we assume a game with *complete* information. Accordingly, an argument game is interpreted as an *extensive games with perfect and complete information*. The players are the opponent and the proponent. An history $h = (\mathrm{turn}_k)_{k=1...n}$ is a dialogue in an argumentation game, and a terminal history is a terminated dialogue. The function assigning to each non-terminal history a player is the player function of the protocol of the argumentation game. A preference relation is defined over terminal histories. The following defines an extensive game with perfect information adapted to our dialectical setting.

Definition 1 *An extensive argumentation game with perfect information is a 4-tuple* $\langle N, H, P, (\succeq_i)\rangle$ *where:*

- $N = \{Opp, Pro\}$ *is a set of arguers, namely the opponent and the proponent;*
- *H is a set of histories (denoted h) which are sequences* $(\mathrm{turn}_k)_{k=1...n}$ *of turns* turn_k*. A history* $(\mathrm{turn}_k)_{k=1...K}$ *is* terminal *if it is infinite or if there is no* turn_{K+1} *such that* $(\mathrm{turn}_k)_{k=1...K+1} \in H$*. The set of terminal histories is denoted by Z;*
- *P is a function that assigns to each non-terminal history a member of N in such a way that the arguers change turns;*
- $\succeq_i$ *is a preference relation on Z for each arguer* $i \in N$*.*

A convenient representation of the argument game tree of our example, interpreted as an extensive game with perfect information, is provided in Figure 2. The set H of histories consists of the (partial) branches of the tree, and the set Z of terminated histories consists of the branches ending with the empty turn $(\emptyset)$.

We adopt the usual convention that if h denotes a history and t turn, then (h,t) denotes the history that results if history h is followed by the turn t . After any nonterminal history h player $P(h)$ chooses a move from the set $M(h) = \{t|(h,t) \in H\}$.

The strategy of an arguer is defined as the specification of the sequence of arguments chosen by the arguer for every history after which it her turn to move (see [5], p. 92).

Argument	A	B	C	D	E
Arguer	*Pro*	*Opp*	*Opp*	*Pro*	*Pro*
Construction chance	$0,6$	$0,7$	$0,4$	$0,3$	$0,6$
$Cost^*_{Pro}\,/\,Ben^*_{Pro}$	$0/0$	$0/0$	$0/0$	$0/0$	$0/0$
$Cost^{Succ(A)}_{Pro}\,/\,Ben^{Succ(A)}_{Pro}$	$0/300000$	$0/0$	$0/0$	$210000/0$	$0/0$
$Cost^{\neg Succ(A)}_{Pro}\,/\,Ben^{\neg Succ(A)}_{Pro}$	$0/0$	$0/0$	$0/0$	$0/0$	$0/0$
$Cost^*_{Opp}\,/\,Ben^*_{Opp}$	$0/0$	$0/0$	$0/0$	$0/0$	$0/0$
$Cost^{Succ(A)}_{Opp}\,/\,Ben^{Succ(A)}_{Opp}$	$2400/0$	$0/0$	$0/0$	$0/0$	$0/0$
$Cost^{\neg Succ(A)}_{Opp}\,/\,Ben^{\neg Succ(A)}_{Opp}$	$0/2400$	$0/0$	$0/0$	$0/0$	$0/0$

Table 1. Flat example's arguments. The construction chance and, the different costs and benefits for arguer i denoted $Cost^*_{Pro}\,/\,Ben^*_{Pro}$, $Cost^{Succ(A)}_i\,/\,Ben^{Succ(A)}_i$ and $Cost^{\neg Succ(A)}_i\,/\,Ben^{\neg Succ(A)}_i$ are explained in Section 4.

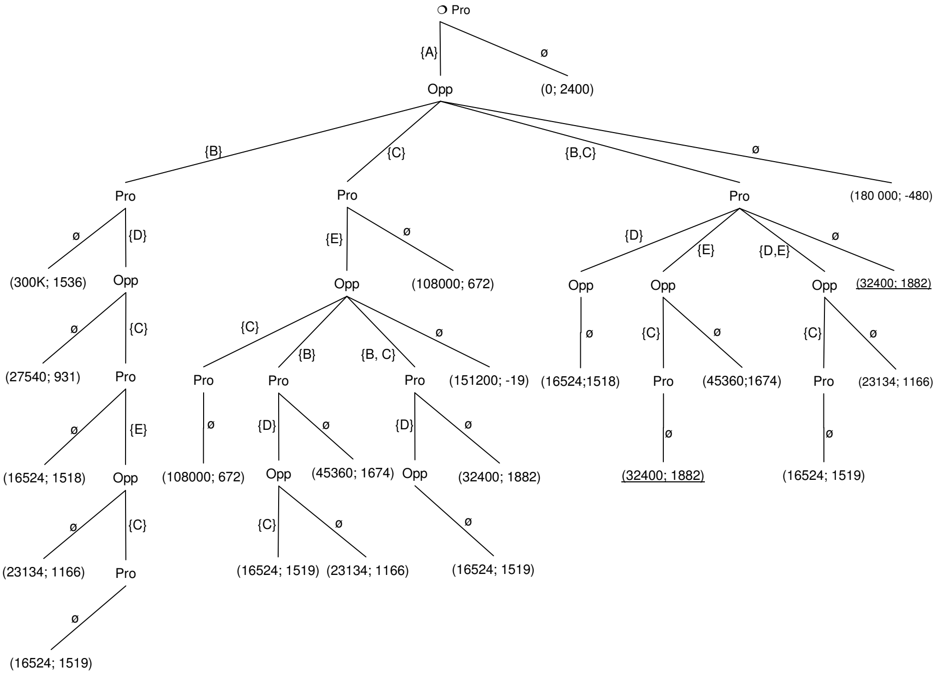

Figure 2. An extensive game tree. The payoffs for the Proponent and Opponent are indicated for each terminal game in terms of expected utility (see Section 4). Underlined payoffs correspond to the perfect equilibria.

Definition 2 *A strategy of arguer $i \in N$ in an extensive argumentation game with perfect information $\langle N, H, P, (\succeq_i)\rangle$ is a function that assigns a move $M(h)$ to each nonterminal history $h \in H - Z$ for which $P(h) = i$.*

For each strategy profile $s = (s_i)_{i \in N}$, the outcome $Out(s)$ is the terminal history that results when each player follows her strategy function.

As [5] (p. 93-97) observe, not all Nash equilibria are plausible in extensive game with perfect information; to identify solutions we need the notions of *subgame* and *subgame perfect equilibrium*. For our dialectical setting, such notions are defined as follows.

Definition 3 *The subgame of the extensive argumentation game with perfect infor-*

mation $\Gamma = \langle N, H, P, (\succeq_i) \rangle$ *that follows the history* h *is the extensive game* $\Gamma(h) = \langle N, H|_h, P|_h, (\succeq_{i|h}) \rangle$*, where*

- $H|_h$ *is the set of sequences* h' *of turns for which* $(h, h') \in H$*, and*
- $P|_h$ *is defined by* $P|_h(h') = P(h, h')$ *for each* $h' \in H|_h$*, and*
- $\succeq_{i|h}$ *is defined by* $h' \succeq_{i|h} h''$ *if and only if* $(h, h') \succeq_i (h, h'')$*.*

The strategy s_i in the subgame $\Gamma(h)$ is denoted $s_i|_h$ while the outcome function of $\Gamma(h)$ is denoted Out_h. Next we provide an adaptation for argument game of the definition of a subgame perfect equilibrium of an extensive game with perfect information:

Definition 4 *A subgame perfect equilibrium of an extensive argumentation game with perfect information* $\Gamma = \langle N, H, P, (\succeq_i) \rangle$ *is a strategy profile* s^* *such that for every nonterminal history* $h \in H - Z$ *for which* $P(h) = i$*,* $i \in \{Opp, Pro\}$*, we have:*

$$Out_h(s^*_{Pro}|_h, s^*_{Opp}|_h) \succeq_{Opp|h} Out_h(s^*_{Pro}|_h, s_{Opp})$$

$$Out_h(s^*_{Pro}|_h, s^*_{Opp}|_h) \succeq_{Pro|h} Out_h(s_{Pro}, s^*_{Opp}|_h)$$

for every s_{Pro} *and* s_{Opp} *in the subgame* $\Gamma(h)$*.*

The preference relation is defined by means of an utility function $EU_i : Out(s) \to \mathbb{R}$ such that $Out(s) \succeq_i Out(s')$ if and only if $EU_i(Out(s)) \geq EU_i(Out(s'))$. The equilibria of our example corresponds to the outcomes $(\{A\}, \{B, C\}, \emptyset)$ and $(\{A\}, \{B, C\}, \{E\}, \{C\}, \emptyset)$ which both have as associated payoff profile $(32\,400, 1\,882)$.

The subgame perfect equilibrium can be compiled by using standard backwards induction (see e.g. [5], p. 99). Briefly, backwards induction means that one starts at a player's final decision nodes to see what a player will do there, and then reasons backwards to tell which action is best for the other player.

In a game in which any strategy leads to the same payoff, game theory is trivially useless. Accordingly, game theory is useful in games in which some strategies lead to different outcomes.

Hereafter, we use the following terminology. An arguer is the *expected winner* if, in the tree with the initial argument as root and for each node containing all replies that are legal according to the 'basic' argument game, the dialogical status of the root favours that player. An arguer is the *expected loser* if he is not the expected winner. For any arguer i, the utility function, which assigns value 1 if arguer i wins and 0 if arguer i loses, is called the *minimal utility function*. Let us finally call a game protocol *sound* and *complete* if it guarantees that the tree as just defined is traversed in every game. In a game with a sound and complete protocol and a minimal utility function, all the strategies of the expected winner would assure her success at winning the game. Any strategy of the expected loser assures her to lose. Hence, in a game with a sound and complete protocol and minimal utility function, game theory is useless. Accordingly, in a game *with* a multi-move protocol which is sound and complete, associated to a minimal utility function, game theory is also useless.

A protocol which is not sound and complete may provide strategies leading to different outcomes. For example, in a protocol in which withdrawals are legal (as in the present paper), game theory allows the expected loser to determine strategies which are not leading to her defeat.

Perhaps more interesting, a game in which the utility function is not minimal may also provide strategies leading to different outcomes (even if the protocol is sound and complete). For example, game theory in any game with a multi-move, sound and complete protocol in association with a utility function compiling costs of moves may help the expected winner to determine the less costly strategy to win. It may be also sometimes better for an arguer to advance fewer arguments than moving more arguments. For instance, the proponent John Partor does not move D because it is too much costly: after the history $(\{A\},\{B,C\})$, the proponent is better off to argue E instead of $\{D,C\}$.

Changing a protocol may involve changing the perfect equilibria of the game. For example, if the protocol in our case were not multi-move, then all the histories of the extensive game tree in Figure 2 implying a turn with more than one argument would not exist, and the outcome $(\{A\},\{B\},\emptyset)$ would become the unique perfect equilibrium. This shows that game theory is a useful method to test the "fairness" of protocols in terms of their completeness and soundness.

4. Preference specification

In this section we provide a specification of the arguers'preferences over outcomes in terms of expected utility.

As is well-known, different options are available to calculate expected utility (EU) in decision theory and there is an extensive debate in the literature. Well-established trends in decision theory state that decisions should correspond to choosing between risky or uncertain prospects by comparing their expected utility values, which are the weighted sums obtained by adding the utility values of outcomes multiplied by their probabilities. This intuition raises in contemporary debate many questions. For example, what are utility numbers referring to? Are they measured using the same value scale adopted under certainty? Again, Is the weighted sum procedure the only one to be considered? Should it be taken for granted that we rely on probability values, or are there alternative constructions? Finally, we can usually follow two well-known versions of decision theory, namely, Subjective Expected Utility Theory in the case of uncertainty, and von Neumann-Morgenstern Theory in the case of risk.

For the sake of simplicity, we adopt the most classical way to calculate *EU* of an act X, which is the sum of the products of the probabilities and utility's values for each outcome, formally, $EU(X) = \sum_{i=1}^{n} Pr(o_i).u(o_i)$ where $o_1,\ldots,o_n$ are the possible (and mutually exclusive) outcomes of X.

In our setting, an arguer i is interested in her expected utility, with regard to the outcome of a strategy profile s noted $EU_i(Out(s))$. The outcome of s, as we know, is a terminal history, namely, the dialogue resulting from s. The evaluation of the dialogue depends on the status of its initial node, which results from whether the adjudicator accepts the arguments constituting the dialogue. If according to the adjudicator's assessment the initial node is *in*, then the proponent of the dialogue has won. If the initial node is *out*, then the opponent has won. For each terminated game associated to strategy profile s, we have two mutually exclusive outcomes: for each arguer, she can win (the initial move is *in*) or lose (the initial move is *out*). In other words, the initial argument A is successful or not. Hence, we have:

$$\begin{aligned} EU_i(Out(s)) = {} & Pr(Succ(A,Out(s))).u_i(Succ(A,Out(s))) \\ & +Pr(\neg Succ(A),Out(s)).u_i(\neg Succ(A,Out(s))) \end{aligned} \tag{1}$$

where $Pr(Succ(A, Out(s)))$ denotes the probability of success of the initial argument A w.r.t. the dialogue $Out(s)$ and $u_i(Succ(A, Out(s)))$ is the utility value of the success of A w.r.t. the dialogue $Out(s)$.

The probability of success of an argument is intended to mean the probability that the argument is accepted as justified given a knowledge base of which the statements are assigned a probability of acceptance by the adjudicator. The method adopted here is similar to that in [8], but it is slightly adapted to the dialectical setting of the present paper: the probability of success of an argument A w.r.t. to an argument game is the probability of success of A w.r.t. the corresponding reply tree. Also, for the sake of simplicity, we assume that no premise of one argument is a conclusion of another argument.[3] The reader is referred to [8] for details and some discussions.

In this setting, the probability of success of an argument depends on two conditions:

1. the probability that the argument's premises are accepted, which we call *construction chance* (the argument will be rejected if the adjudicator refuses to accept one of the argument's premises);
2. the probability that the argument has no valid counterargument, namely no counterargument is able to attack it successfully, which we call the *security chance* (the argument will be rejected if the adjudicator's acceptances imply that there is a valid attacker of the argument).

The construction chance of an argument A, denoted by $Pr(Con(A))$, is given by the probability that the adjudicator accepts all premises $q_1, \ldots q_n$ of the argument, that is, by the probability of $q_1^a, \wedge \ldots \wedge q_n^a$, where q_i^a denotes the acceptance of q_i. On the assumption that the premises of an argument are mutually (statistically) independent, $Pr(Con(A))$ is the product of the probability of acceptance of all premises in the argument:

$$Pr(Con(A)) = Pr(q_1^a \wedge \ldots \wedge q_n^a) = Pr(q_1^a) \times \ldots \times Pr(q_n^a) \qquad (2)$$

Let us now add the second condition of probability of success of an argument, namely that of not having a successful counterargument. The basic idea is that the chance of success of an argument is diminished to the extent that one of its attackers is going to be successful. Thus, generally, the probability of success of an argument A is diminished by considering the chances of success of all of its counterarguments, along all possible branches of the reply tree of which the root is given by A. In the following we shall first consider how to compute the security chance along one line (i.e. a branch) of the reply tree, and then how to compute security chance along multiple lines, that is, in a reply tree.

We first characterise the security chance of an argument A_i in a branch $D_n =< A_1, \ldots, A_n >$, intended as the probability that A_i can really exercise its intended function in the branch.

Since each argument, according to the game protocol, is defeated (and thus prevented from being successful) by its successor, the probability of success of an argument does not only depend on its construction chance, but also on the chance that the argument's successor fails to be successful. Since both these elements need to be present, we have to deal with the probability of a conjunction of the construction of the argument and the failure of its attacker. We therefore define $Succ(A_i, D_n)$ as $Con(A_i) \wedge \neg Succ(A_{i+1}, D_n)$. Since equivalent statements are equally probable we have:

[3] In any case, this assumption is reasonable since if some argument's premise is the conclusion of another argument, the two arguments should ideally be combined by making the second argument a subargument of the first, thus turning the premise of the first into an intermediate conclusion.

$$Pr(Succ(A_i,D_n)) = Pr(Con(A_i)).[1 - Pr(Succ(A_{i+1},D_n)|Con(A_i))] \tag{3}$$

The chance of success of the last argument A_n of D_n is given by its chance of construction: $Pr(Succ(A_n,D_n)) = Pr(Con(A_n))$.

We need to consider the probability of success of an argument taking into account that the argument can have more than one counterargument, and that the probabilities of success of such counterarguments somehow need to be added up (since it is sufficient that one counterargument is successful for the argument to fail). Consider an arbitrary argument A in a reply tree $\tau = \langle \tau_1, \ldots, \tau_k \rangle$ having counterarguments (children) $A_1, \ldots, A_k$. The probability that any of these counterarguments A_j is successful is the probability of the disjunction $\bigvee_{j=1}^{j=k} Succ(A_j, \tau_j)$. Then the probability that A is successful is given by the probability that A is constructed times the probability that no such counterargument is successful, that is, that the above disjunction is false. Formally:

$$\begin{aligned} Pr(Succ(A,\tau)) &= Pr(Con(A) \wedge [\neg(\textstyle\bigvee_{j=1}^{j=k} Succ(A_j,\tau_j))]) \\ &= Pr(Con(A)).Pr(\textstyle\bigwedge_{j=1}^{j=k} \neg Succ(A_j,\tau_j)|Con(A)) \\ &= Pr(Con(A)).\textstyle\prod_{j=1}^{j=k}\{Pr(\neg Succ(A_j,\tau_j)|Con(A) \textstyle\bigwedge_{i=1}^{i=j-1} \neg Succ(A_i,\tau_i))\} \end{aligned} \tag{4}$$

For example, the security chance of argument A along the dialogue $(\{A\},\{B,C\},\{D,E\},\{C\})$ is (assuming that the premises of the arguments are statistically independent):

$$\begin{aligned} & Pr(Succ(A,(\{A\},\{B,C\},\{D,E\},\{C\}))) \\ &\quad = Pr(Con(A)).\{Pr(\neg Succ(B,(\{B\},\{D\}))|Con(A) \wedge \neg Succ(C,(\{C\},\{E\},\{C\}))) \\ &\qquad\qquad .Pr(\neg Succ(C,(\{C\},\{E\},\{C\}))|Con(A) \wedge \neg Succ(B,(\{B\},\{D\})))\} \\ &\quad = Pr(Con(A)).Pr(\neg Succ(B,(\{B\},\{D\}))).Pr(\neg Succ(C,(\{C\},\{E\},\{C\}))) \\ &\quad = Pr(Con(A)).[1 - Pr(Succ(B,(\{B\},\{D\})))].[1 - Pr(Succ(C,(\{C\},\{E\},\{C\})))] \\ &\quad = 0,1836 \end{aligned} \tag{5}$$

Interestingly, one can demonstrate by induction that the probability of success $Pr(Succ(A_1,D_n))$ of the initial claim A_1 w.r.t. a single argument game $D_n =< t_1,\ldots,t_n >$ is bounded in such a way that $Pr(Succ(A_1,t_i)) \geq Pr(Succ(A_1,t_j))$ if i is odd, and $Pr(Succ(A_1,t_i)) \leq Pr(Succ(A_1,t_j))$ if i is even, where $j \geq i$. Hence, the probabilities $Pr(Succ(A_1,t_1))$ and $Pr(Succ(A_1,t_2))$ are respectively the highest bound and lowest bound of $Pr(Succ(A_1,t_j))$. Within these bounds, $Pr(Succ(A_1,t_j))$ oscillates at every move, with the maximum amplitude that decreases with the length of the dialogue.

Next, we focus on the utility values $u_i(Succ(A,Out(s)))$ and $u_i(\neg Succ(A,Out(s)))$ to incorporate costs and benefits of moves. In general, we can distinguish between fixed costs/benefits and costs/benefits dependant upon success. The former ones, in particular, capture costs/benefits independent of the success of the player: for example, some trial expenses for an arguer i are a fixed cost, since they applies to i independently of the fact that i wins or loses. The utility value $u_i(Succ(A,Out(s)))$ is the sum (over the argument A_k member of $Out(s)$) of the fixed benefits $Ben_i^*(A_k)$ minus the fixed costs $Cost_i^*(A_k)$, plus the sum of benefits $Ben_i^{Succ(A)}(A_k)$ dependant of the success of A minus the costs $Cost_i^{Succ(A)}(A_k)$ dependant of the success of A. The utility value $u_i(\neg Succ(A,Out(s)))$ is the sum of the fixed benefits $Ben_i^*(A_k)$ minus the fixed costs $Cost_i^*(A_k)$, plus the sum of benefits $Ben_i^{\neg Succ(A)}(A_k)$ dependant of the unsuccess of A minus the costs $Cost_i^{\neg Succ(A)}(A_k)$ dependant of the unsuccess of A. Let the set $Arg(out(s)) = \{A_1,\ldots,A_K\}$ of arguments constituting $out(s)$, we have formally:

$$u_i(Succ(A,out(s))) = \sum_{k=0}^{k=K}(Ben_i^*(A_k)+Ben_i^{Succ(A)}(A_k)) - (Cost_i^*(A_k)+Cost_i^{Succ(A)}(A_k))$$
$$u_i(\neg Succ(A,out(s))) = \sum_{k=0}^{k=K}(Ben_i^*(A_k)+Ben_i^{\neg Succ(A)}(A_k)) - (Cost_i^*(A_k)+Cost_i^{\neg Succ(A)}(A_k)) \quad (6)$$

For example, consider the outcome $Out^* = (\{A\}, \{B,C\},\{E\},\{C\},\emptyset)$, we have $Arg(out^*) = \{A,B,C,E\}$, and the proponent's expected utility of Out^* is $EU_{Pro}(Out^*) = 0,1836 \times (300000 - 210000) + (1 - 0,1836) \times 0$, that is, 32400. For the sake of simplicity, we have discarded fixed costs and benefits in our example.

As suggested at the end of Section 3, the perfect equilibrium can be compiled using backwards induction (because the reasoning works backwards from outcomes to present decision problems): a player asks herself which of the available final outcomes brings her the highest utility, and chooses the action that starts the chain leading to this outcome.

5. Related work

This paper is inspired by [9] in which argumentation is modelled as a game where the payoffs are measured in terms of the probability that the claimed conclusion is, or is not, defeasibly provable, given a history of arguments that have actually been exchanged, and given the probability of acceptance of the factual premises. The probability of a conclusion is calculated using Defeasible Logic, in combination with standard probability calculus. How does [9]'s model compare to the present approach? First, [9]'s game can be called a *theory building game*: during such a game the players jointly build a logical theory by exchanges of arguments, and the outcome of a game is determined by checking whether the topic statement is implied by the end state according to a particular logic. Instead we formulate our ideas in terms of [6, 7]'s notion of the *dialogical status* of a dialogue move. In doing so, we abstract on the underlying logic and structure of arguments, to make our ideas as widely applicable as possible. Second, we retain the idea of uncertainty about statement acceptance, but instead of the probability of success of an argument, [9] proposes to compute the probability that the topic is defeasibly justifiable, where the probability of success of an argument plays no role. By considering the latter probability, the dialogue tree is tracked and that may be a decisive advantage for further investigation in the field of argumentation. Third, the utility function in [9] is reduced to the probability of provability of the claim, whereas the utility function of the present paper account also for costs and benefits of strategies and is thus more fine-grained. Fourth, our protocol is multi-move: this allowed us to illustrate that game theory permits to account for cases in which an arguer is better off to advance few arguments in row.

Another work coupling game theory and argumentation is [1] in which the argumentation techniques of [4] and the game theory approach of [2] are integrated to reach agreement by proposing a trade-off in terms of allocation of numerical utilities representing the importance disputants' place on disputed issues. Roughly, the procedure consists of the following: first, the dialogue techniques are used as an attempt to resolve any existing conflicts; second, the issues which are not resolved are the inputs to a compensation/trade-offs process in order to facilitate resolution of the dispute. If the result of the compensation/trade-offs process is not acceptable by the parties, then they return to the the first step and repeat the whole process recursively until either an agreement or a stalemate is reached. [1]'system is meant to be used in *mediation* procedure setting instead of *litigation*. Litigation causes an argument to be discussed in a law court so that a judgment can be made which must be accepted by both sides. Instead, mediation is a process by which the participants, together with the assistance of neutral third party, isolate disputed issues in order to reach an agreement. Accordingly, [2]'s game theory investigation is not used for the same aim and cannot apply to litigation. For example, if only one issue needs to be resolved, then suggesting a trade-off is not possible. In the present paper, we are not interested in reaching an agreement by proposing a trade-off in terms of allocation of utilities representing the importance disputants' place on disputed issues. We aim at determining the optimal strategy of an arguer in dialogue games for argumentation, i.e. an arguer's sequence of moves which optimises her expected utility.

6. Conclusion

The dialectical setting of [6, 7] has been interpreted in game-theoretical terms. This interpretation allowed us to straightforwardly apply game theory, and optimal strategies have been determined accordingly. A specification of preferences over outcomes has been provided in terms of expected utility combining the probability of success of arguments, costs and benefits of arguments. Doing so, we have illustrated that game theory is useful to illuminate diverse aspects of argumentation frameworks.

Future work will focus on assumptions ranging from the specification of the utility function (e.g. assuming non-independent premises) to the game theory modelling (e.g. assuming game with incomplete information, ordinal treatment of subjective utilities). Also, other types of dialogue as negotiation or persuasion could be integrated to the adjudication. Finally, the approach can be implemented into 'argument assistance systems' which offer advice and reasons for its advice, to engage into a legal dispute and to choose the optimal strategies.

References

[1] E. Bellucci, A. R. Lodder, and J. Zeleznikow. Integrating artificial intelligence, argumentation and game theory to develop an online dispute resolution environment. In *ICTAI*, pages 749–754. IEEE Computer Society, 2004.

[2] E. Bellucci and J. Zeleznikow. Representations for decision making support in negotiation. *ournal of Decision Support*, 10(3-4):449–479, 2001.

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

[4] A. R. Lodder. *DiaLaw - on legal justification and dialogical models of argumentation*. luwer Academic Publishers, Law and Philosophy Library, Volume 42, Dordrecht, 1999.

[5] M. J. Osborne and A. Rubinstein. *A Course in Game Theory.* MIT Press, 1999.

[6] H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese, special issue on New Perspectives in Dialogical Logic*, (127):187–219, 2001.

[7] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *J. Log. and Comput.*, 15(6):1009–1040, 2005.

[8] R. Riveret, N. Rotolo, G. Sartor, H. Prakken, and B. Roth. Success chances in argument games: A probabilistic approach to legal disputes. In *Jurix 2007*, To appear, 2007. IOS Press.

[9] B. Roth, R. Riveret, A. Rotolo, and G. Governatori. Strategic argumentation: a game theoretical investigation. In *ICAIL '07: Proceedings of the 11th International Conference on Artificial Intelligence and Law*, pages 81–90, New York, NY, USA, 2007. ACM Press.