Two Faces of Strategic Argumentation in the Law

Guido GOVERNATORI a, Francesco OLIVIERI b, Antonino ROTOLO c, Simone SCANNAPIECO b, Giovanni SARTOR d

a NICTA Queensland Research Laboratory, Australia
b Department of Computer Science, University of Verona, Italy
c CIRSFID, University of Bologna, Italy
d CIRSFID, University of Bologna, and EUI, Italy

Abstract. In strategic argumentation players exchange arguments to prove or reject a claim. This paper discusses and reports on research about two basic issues regarding the game-theoretic understanding of strategic argumentation games in the law: whether such games can be reasonably modelled as zero-sum games and as games with complete information.

Keywords. Strategic argumentation, Zero-sum games, games with incomplete information

1. Introduction

In the most typical and simplest forms of strategic argumentation, two players exchange arguments in a dialogue game: the proponent (hereafter Pro) has the aim to prove a claim, and the opponent (hereafter Opp) presents counterarguments to the moves of Pro.

Over the years many dialogue games for argumentation have been proposed (see among others [22, chaps. 13 - 14]) to study questions such as which conclusions are justified, or how procedures for debate and conflict resolution should be structured to arrive at a fair and just outcome. However, despite the game-like character of arguments and debates, game-theoretic investigations of argumentation are still rare in the AI argumentation literature and in the game theory one as well (an exception in this second perspective is [7]). Exceptions addressing in AI the game-theoretic aspects of strategic argumentation are [23,20,14,21,13], which, however, very often assume to work under ideal conditions that make the resulting frameworks of little use in the legal domain. This paper discusses and reports on research in regard to two basic issues:

- whether strategic argumentation games in law can be reasonably modelled as zero-sum games;

---

1 Corresponding Author: antonino.rotolo@unibo.it. The author is partially supported by the UNIBO project FARB 2012 Mortality Salience, Legal and Social Compliance, and Economic Behaviour: Theoretical Models and Experimental Methods.

2 NICTA is funded by the Australian Government and the Australian Research Council through the ICT Centre of Excellence program.
whether strategic argumentation games in law can be reasonably modelled as games with complete information.

The layout of the paper is the following. Section 2 presents some basic logical and dialectical features of strategic argumentation games. Sections 3 and 4 offer some considerations of why strategic argumentation games in law are not, respectively, zero-sum games and games with complete information. In regard to the latter question, in particular, we report on some recent results (proposed by some of the authors of this paper) and briefly suggest how to extend them.

2. Strategic Argumentation

We assume that strategic argumentation games be based on Dung [5]’s abstract argumentation frameworks. Let $AF = (\mathcal{A}, \triangleright)$ be a Dung’s framework, where $\mathcal{A}$ is as usual a finite non-empty set of arguments and $\triangleright$ is binary attack relation on $\mathcal{A}$. A strategic argumentation game based on any Dung’s framework $AF$ has the the form of a dialogue game, which involves a sequence of interactions between two players, the Proponent Pro and the Opponent Opp. The content of the dispute being that Pro attempts to assess the validity of a particular thesis, whereas Opp attacks Pro’s claims in order to refute such thesis. The challenge between the parties is formalised by means of argument exchange using arguments and attacks encoded in $AF$.

Given any $AF$, any strategic argumentation game is based on the following protocol, which adopts the the dialectical setting for disputes (in the context of persuasion dialogues) presented in [17,18]:

- 2-person argumentation game is played by two players Pro and Opp;
- A move in an argumentation game is a tuple $(pl, id : a)$ where $pl$ is the player of the move, $id$ is the move identifier (a natural number), $a$ is the set of arguments moved in $id$ such that
  - if $a = \emptyset$, then the move is a withdrawal;
  - if $id = 1$ (i.e., the first move), then $a$ is either $\emptyset$ (withdrawal: no debate takes place) or a singleton (only one argument is played);
  - if $id > 1$, then $a = \{A_1, \ldots, A_n\}$, $0 \leq n \leq m$ ($m$ being the number of arguments in the set $\mathcal{A}$ of $AF$), such that all arguments played in the previous move are defeated by an argument in $a$;
- Player Pro does not repeat moves;
- If a move is legal then it satisfies all preceding conditions. A withdrawal move is always legal.

With regard to argumentation games, as done in [23], we make the following assumptions:

1. An argumentation game terminates if a player withdraws. Since the set of arguments in any $AF$ is finite, then each game terminates, because Pro may not repeat moves.
2. Each argumentation game induces a reply tree, which consists of the argument moves as nodes and their target relations as links.
3. Reply trees can be labeled as follows: arguments in a node are in iff all arguments in its children are out; and arguments in a node are out iff it has a child with arguments that are in [3]. (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.) 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 [5].

4. An argument move $M$ in a reply tree $T$ favours $\text{Pro}$ if all arguments in $M$ is in; otherwise $M$ favours $\text{Opp}$.

5. A game is won by a player if at termination the initial move favours the player.

3. Zero-sum vs Non-zero-sum Argumentation Games

One may have good reasons to assume that strategic argumentation games are strictly competitive, i.e., informally, that they are games where each strategy profile that is better for one player is worse for the other player. This formalises the intuition that the interests of both players are diametrically opposed [15]. Since legal disputes—and, more generally, persuasion dialogues (see [17])—have an adversarial form and their purpose is establishing the winner and the loser between the players, assuming to work with such a type of games can be viewed as natural option. More specifically, if such games are strictly competitive and preferences are expressed by utility functions, then, for any strategy profile the sum of payoffs of the players gives 0.

Recent works such as [13] have in fact taken this general view, which offers several technical advantages (some standard ones being, e.g., that strategies in Nash equilibria are maximinimizer and yield the same payoffs).

However, consider the following fictional but realistic legal scenario:

Example 1 (Flat example [23]) The proponent $\text{Pro}$, 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 $\text{Opp}$, 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 require 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 for, 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 formalised by a notary deed. Alternatively, Rex may present the argument $B$ that the market value of the flat is 120000 euros and so the rental agreement is valid, whereas John may reply with $D$ saying that he will pay within
The example has been reconstructed in [23] according to the following assumptions:

- argumentation dialogues in law are modelled as extensive games (with complete information) to provide an explicit account of the sequential structure of argumentation moves;
- optimal strategies are determined by preferences over outcomes of the disputes: payoffs are calculated as standard expected utilities of strategies by combining the probability of success of arguments with their associated costs and benefits;
- the cost of each argument $A$ is in general independent from the dialectical status of $A$ (in other words, players will incur in costs independently of whether the played argument succeed);
- the role of an adjudicator is highlighted (like in adjudication dialogues [19]) in such a way that the probability of success of arguments is expressed by assuming a probability distribution with respect to the adjudicator’s acceptance of the parties’ statements: this distribution determines the probability of the argument’s success, which is to be established on the basis also of the probabilities of success of its counterarguments.

For each strategy profile $s$, the outcome $Out(s)$ is the terminal history that results when each player follows her strategy function. Then, if the probability of success of arguments are, for example, as in Table 1, the payoffs are calculated as follows [23].

$$EU_i(Out(s)) = Pr(Succ(A, Out(s))).u_i(Succ(A, Out(s))) + Pr(\neg Succ(A, Out(s))).u_i(\neg Succ(A, Out(s)))$$

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 $Out(s)$.

On the basis of the above analysis, the reply tree for Example 1 was reconstructed in [23] as in Figure 1, which trivially shows that in this case the sums of payoffs in strategy profiles do not give 0. Figure 1 also shows (at the right-hand side of the tree) that there exists a strategy profile corresponding to a subgame perfect equilibrium [23, 15] where the Proponent’s move is withdrawal. This second observation is interesting because it shows that any attempt of abstracting in this scenario from utility functions cannot faithfully lead in any case to a strictly competitive game.

Hence the following fact can be stated:

**Fact 1 (Zero-sum vs Non-zero-sum games)** If the preferences of players are expressed in strategic argumentation games by utility functions and the costs are independent from the dialectical status of the arguments to which costs are associated, then strategic argumentation games are non-zero-sum games.

<table>
<thead>
<tr>
<th>Argument</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arguer</td>
<td>Pro</td>
<td>Opp</td>
<td>Opp</td>
<td>Pro</td>
<td>Pro</td>
</tr>
<tr>
<td>Probability</td>
<td>0.6</td>
<td>0.7</td>
<td>0.4</td>
<td>0.3</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Table 1. Probability of success in flat example

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.
Figure 1. An extensive game tree illustrating Example 1. Branches labelled with $\emptyset$ corresponds to withdrawal moves, i.e., moves where the player quits. The payoffs for the Proponent Pro and Opponent Opp are indicated for each terminal game in terms of expected utility. Underlined payoffs correspond to the perfect equilibria.

Indeed, Example 1 shows that this result (that strategic games are not necessarily zero-sum games) intuitively depends on the possibility of parties’ strategic withdrawals: when the costs of moves is independent of their dialectical success (think about the costs of experts’ consulting services), withdrawing can be rational and even optimal.

4. Complete vs Incomplete Information in Argumentation Games

While several investigations in the AI scholarship on formal argumentation do not assume any special conditions on the players’ knowledge in argumentation games (see, e.g., [8,4,16,18,2]), most, if not all existing works explicitly investigating in AI the game-theoretic aspects of strategic argumentation, such as [20,14,21,13,23], assume instead that argumentation games have complete information, i.e., that the structure of the argumentation game (typically, the set of all possible arguments) is common knowledge among the players.

This is another oversimplification, as in many real-life contexts (such as in legal disputes) players do not know the entire structure of the argumentation game: in fact, each of them does not know what arguments her opponent will employ. Consider the following example:

Example 2 ([24]) Suppose the following witness examination takes place in a court:
Pro, 1: “You killed the victim.”
Opp, 2: “I did not commit murder! There is no evidence!”
Pro, 3: “There is evidence. We found your ID card near the scene.”
Opp, 4: “It’s not evidence! I had my ID card stolen!”
Pro, 5: “It is you who killed the victim. Only you were near the scene at the time of the murder.”
Opp, 6: “I didn’t go there. I was at facility A at that time.”
Pro, 7: “At facility A? Then, it’s impossible to have had your ID card stolen since facility A does not allow a person to enter without an ID card.”

The peculiarity of this game is that the exchange of arguments reflects an asymmetry of information between the players: first, each player does not know the other player’s knowledge, thus she cannot predict which arguments are attacked and which counterarguments are employed for attacking the arguments; second, the private information disclosed by a party could be eventually used by the adversary to construct and play justified counterarguments: the argument played by Pro at move 7 attacks the one played by Opp at 4, but only after Opp has played the argument at move 6. Hence, the attack at 7 of the proponent is made possible only when the opponent discloses some private information with the move at 6.

Notice that such an asymmetry of information is quite common in legal disputes and it does not only regard factual knowledge (as in the example above) but also normative knowledge: if Pro (Opp) knows more normative arguments (e.g., based on statutory or case law), she could be advantaged by that fact, but she must be careful in how and when playing such arguments because they could be used by Opp (Pro) to attack her in combination with other argument that only Opp (Pro) knows.

4.1. Computational Issues in Strategic Argumentation with Incomplete Information

Handling argumentation games with incomplete information can be computationally hard. Recent results such as [11, 10] have considered in particular the case of argumentation frameworks with structured arguments, i.e., where the (internal) logical structure of arguments can be specified in such a way that arguments are logical inference trees. The logic used to build arguments is Defeasible Logic [1]. The knowledge base in this logic consists of defeasible theories such as $D = (F, R, >)$, where $F$ is a set of indisputable facts, $R$ is a set of rules, and $>$ is binary superiority relation over rules in $R$.

In this context, strategic argumentation games can be basically designed as in Section 2. However, since arguments have an internal logical structure resulting from chaining rules of Defeasible Logic in order to prove literals, argument moves have here the form of defeasible theories and derivations of the desired conclusions.

When argumentation games have incomplete information, given an arbitrary large set of rules (or arguments), players are initially equipped with different logical theories which constitute their private knowledge, being unknown by the opposite party: in other words, each player does not know what arguments or rules are taken to be valid by the opposite party in the game. Hence, a player may build an argument supporting her claim by using some of her private knowledge; in turn, the other party may then re-use such and others rules (again from her own private knowledge) to construct a new argument defeating the previously constructed argument. In other words, the set $R$ of rules that
are used to build arguments is partitioned into three subsets: a set $R_{\text{Com}}$ known by both players and two subsets $R_{\text{Pro}}$ and $R_{\text{Opp}}$ corresponding, respectively, to Pro’s and Opp’s private knowledge. Consider a setting where $F = \{a, d, f\}$ is the known set of facts, $R_{\text{Com}} = \emptyset$, and the players have the following rules:

$$R_{\text{Pro}} = \{a \Rightarrow b, \ d \Rightarrow c, \ c \Rightarrow b\} \quad R_{\text{Opp}} = \{c \Rightarrow e, \ e, f \Rightarrow \neg b\}.$$  

If Pro’s intent is to prove $b$ and she plays $\{a \Rightarrow b\}$, then Pro wins the game. If Pro plays $\{d \Rightarrow c, \ c \Rightarrow b\}$ (or even $R_{\text{Pro}}$), Opp will instead succeed. Here, a minimal subset of $R_{\text{Pro}}$ is successful. The situation can be reversed for Pro. Replace the sets of private rules with

$$R_{\text{Pro}} = \{a \Rightarrow b, \ d \Rightarrow \neg c\} \quad R_{\text{Opp}} = \{d, c \Rightarrow \neg b, \ f \Rightarrow c\}.$$  

In this second case, the move $\{a \Rightarrow b\}$ is not successful for Pro, while playing with the whole $R_{\text{Pro}}$ ensures victory.

The state of the game at move $i$ is denoted by a defeasible theory $D_i = (F, R_{\text{Com}}, >)$ and by two sets $R_{\text{Pro}}^i$ and $R_{\text{Opp}}^i$. $R_{\text{Com}}^i$ is the set of rules known by both players at $i$ (which may be empty when $i = 0$), and $R_{\text{Pro}}^i$ ($R_{\text{Opp}}^i$) is the private knowledge of Pro (Opp) at $i$.

The assumptions in [11, 10] are that

1. each player is informed about the restriction of $>$ to the rules that she knows,
2. $D_i$ is coherent and consistent for each $i$, i.e., there is no literal $p$ such that: (i) $p$ is at the same time defeasibly proved and refuted in $D_i$, and (ii) both $p$ and $\neg p$ are defeasibly proved in $D_i$.

The parties start the game by choosing the critical literal $l$ to discuss about and run the game as follows.

**Definition 1** Let $R_{\text{Com}}^1$, $R_{\text{Pro}}^1$ and $R_{\text{Opp}}^1$ be respectively the common knowledge, and the private knowledge of Pro and Opp at the beginning of the game, and $D^1 = (F, R_{\text{Com}}^1, >)$. If $D^1 \models l$, then Opp starts the game. Otherwise, Pro does so.

At move $i$, if Pro plays $R^i$, then:

- $D_i^{i-1} \models \neg l$;
- $R_i \subseteq R_{\text{Pro}}^{i-1}$;
- $D_i = (F, R_{\text{Com}}^i, >)$;
- $R_{\text{Pro}}^i = R_{\text{Pro}}^{i-1} \setminus R^i$; $R_{\text{Opp}}^i = R_{\text{Opp}}^{i-1}$, and $R_{\text{Com}}^i = R_{\text{Com}}^{i-1} \cup R^i$;
- $D_i \models l$.

At move $i$, if Opp plays $R^i$, then:

- $D_i^{i-1} \models l$;
- $R_i \subseteq R_{\text{Opp}}^{i-1}$;
- $D_i = (F, R_{\text{Com}}^i, >)$;
- $R_{\text{Pro}}^i = R_{\text{Pro}}^{i-1} \setminus R^i$; $R_{\text{Opp}}^i = R_{\text{Opp}}^{i-1}$, and $R_{\text{Com}}^i = R_{\text{Com}}^{i-1} \cup R^i$;
- $D_i \models \neg l$.

---

[1] Defeasible Logic uses tagged literals to denote the modes of derivation (strict or defeasible) and whether the literal is provable or not provable in a given theory. We omit these details to simplify the presentation.
In this context, we can formulate the following decision problem:

**Strategic Argumentation Problem**

**Pro’s instance for move $i$:** Let $l$ be the critical literal, $R_{Pro}^{i-1}$ be the set of the private rules of Pro, and $D_{Pro}^{i-1}$ be such that $D_{Pro}^{i-1} \vdash \neg l$.

**Question:** Is there a subset $R^i$ of $R_{Pro}^{i-1}$ such that $D^i \vdash l$?

**Opp’s instance for move $i$:** Let $l$ be the critical literal, $R_{Opp}^{i-1}$ be the set of the private rules of Opp, and $D_{Opp}^{i-1}$ be such that $D_{Opp}^{i-1} \vdash l$.

**Question:** Is there a subset $R^i$ of $R_{Opp}^{i-1}$ such that $D^i \vdash \neg l$?

In other words, the problem above amounts to deciding what set of rules to successfully play at any given move.

[11] has considered standard propositional Defeasible Logic [1] while in [10] the analysis is extended by considering one variant of Defeasible Logic characterising one of the most popular argumentation semantics proposed by Dung [5], i.e., grounded semantics. In both cases (with standard Defeasible Logic as well as with Defeasible Logic under grounded semantics), the following negative result holds:

**Theorem 1** The Strategic Argumentation Problem is NP-complete under defeasible and grounded semantics.

The above analysis can be generalised in order to investigate Dung’s abstract argumentation frameworks. First of all, it is worth noting that characterising argumentation games with incomplete information in Dung’s perspective can amount to adopting one of the following two options (or a combination of both):

**Definition 2 (AF with incomplete information (1))** An abstract argumentation framework with incomplete information $AF$ is a structure $(\mathcal{A}, \succsim)$ where

- $\mathcal{A} = \mathcal{A}_{Com} \cup \mathcal{A}_{Pro} \cup \mathcal{A}_{Opp}$ is a finite non-empty set of arguments consisting of arguments that are common knowledge ($\mathcal{A}_{Com}$), and arguments that belong to the private knowledge of players ($\mathcal{A}_{Pro}$ and $\mathcal{A}_{Opp}$);
- $\succsim$ is binary attack relation on $\mathcal{A}$.

**Definition 3 (AF with incomplete information (2))** An abstract argumentation framework with incomplete information $AF$ is a structure $(\mathcal{A}, \succsim)$ where

- $\mathcal{A}$ is a finite non-empty set of arguments consisting of arguments that are common knowledge;
- $\succsim = \succsim_{Com} \cup \succsim_{Pro} \cup \succsim_{Opp}$ is binary attack relation on $\mathcal{A}$ consisting of attacks that are common knowledge ($\succsim_{Com}$), and attacks that belong to the private knowledge of players ($\succsim_{Pro}$ and $\succsim_{Opp}$).

Definition 3 can offer an interesting reconstruction of Example 2: indeed, one may argue that Opp is unaware of the fact that the argument she played at 6 attacks another argument she played before, i.e., the one played at move 4.

However, from the computational point of view we can just consider Definition 2, since any NP-completeness results for that case can be trivially extended to the case where the assumption of incomplete information regards the attack relation.
If we assume Definition 2, Theorem 1 is expected to hold in general. Indeed, as observed in [11] the Strategic Argumentation Problem is structurally similar to another NP-complete problem, the Restoring Sociality Problem [12], which is in turn mapped into a known problem, Knapsack Problem [6, Problem MP9]. This mapping between problems does not essentially depend on the use of Defeasible Logic: it is in fact a routine exercise to reframe Definition 1 and replace literal $l$ with a critical argument and sets of rules with sets of arguments. The complexity of the problem thus depends on the fact that establishing if a move is successful can require to explore all subsets of $\mathcal{A}$.

The reader with a little familiarity with these questions would appreciate that the complexity of the problem depends on the asymmetry of information among the parties of legal discourse. If the set of arguments is arbitrarily large but finite, the problem is decidable, but, still, each party does not know what the other party knows, and this fact can make intractable the problem of understanding if a move is successful. This is relevant even when one of the party knows the entire set of arguments: for example, this party could not aware of that or, if she is aware, assuming that she cannot win (something that she knows by computing the set of successful arguments), she could anyway decide to play counting on the fact that the opposite party does not know all arguments.

Finally, notice that, in strict game-theoretic terms, one of the simplest ways of analysing argumentation games of incomplete information is to frame them as Bayesian extensive games with observable actions [15, chap. 12]: this is possible because every player observes the argumentative move of the other player and uncertainty only derives from an initial move of Chance that distributes (payoff-relevant) private information among the players corresponding to logical theories: hence, chance selects types for the players by assigning to them possibly different theories from the set of all possible theories constructible from a given language. If this hypothesis is correct, notice that Bayesian extensive games with observable actions allow to simply extend the argumentation models proposed, e.g., in [23,13]. Despite this fact, however, complexity results for Bayesian games are far from encouraging (see [9] for games of strategy). This problem is left to future research.

5. Summary

In this paper we discussed two basic issues concerning the game-theoretic nature of strategic argumentation in the law: (a) whether strategic argumentation games in law can be reasonably modelled as zero-sum games; (b) whether strategic argumentation games in law can be reasonably modelled as games with complete information.

As regards the first issue, using the framework developed in [23], we showed that whenever the preferences of players are expressed by utility functions and costs are independent from the dialectical status of the arguments to which costs are associated, then strategic argumentation games are non-zero-sum games. As regards the second issue, we argued on the basis of [11,10] that the assumption of complete information is a simplistic idealisation in the law. We recalled, however, that moving to games with incomplete information may be computationally costly, thus undermining any efficient model for strategic argumentation for the law.
References


