Strategic Argumentation is NP-Complete

G. Governatori[1,2,5] **F. Olivieri**[1,2,3] **S. Scannapieco**[1,2,3] **A. Rotolo**[4] **M. Cristani**[3]

Abstract. We study the complexity of the Strategic Argumentation Problem for 2-player dialogue games where a player should decide what move (set of rules) to play at each turn in order to prove (disprove) a given thesis. We show that this is an NP-complete problem.

1 Introduction and Motivation

In the most typical forms of strategic argumentation, two players exchange arguments in a dialogue game: in the simplest case, a proponent (hereafter Pr) has the aim to prove a claim, and an opponent (hereafter Op) presents counterarguments to the moves of Pr. Almost all the AI literature on the strategic aspects of argumentation works with argument games with *complete information*, i.e., dialogues where the structure of the game is common knowledge among the players. Consider, however, the following example due to [17]:

Pr_0 : "You killed the victim."
Op_1 : "I did not commit murder! There is no evidence!"
Pr_1 : "There is evidence. We found your ID card near the scene."
Op_2 : "It's not evidence! I had my ID card stolen!"
Pr_2 : "It is you who killed the victim. Only you were near the scene at the time of the murder."
Op_3 : "I didn't go there. I was at facility A at that time."
Pr_3 : "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 argument 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: Pr_3 attacks Op_2, but only when Op_3 is given. Hence, the attack Pr_3 of the proponent is made possible only when the opponent discloses some private information with the move Op_3.

In scenarios with incomplete information, parties have different logical theories. Not knowing the other party's arguments implies that there is no general way to determine which argument is the most likely to succeed. A party does not know whether an argument is not attacked by arguments from the opponent or whether it allows counterarguments based on it or parts of it (i.e., subarguments). The example above reveals that the internal logical structure of arguments plays a key role, which cannot be overlooked, in strategic argumentation.

In this paper, we explore the computational cost of argument games with incomplete information where the (internal) logical structure of arguments is considered. We study strategic argumentation in proof-theoretic settings, such as in those developed in [4, 13, 19] where arguments are defined as logical inference trees, and exchanging arguments means exchanging logical theories proving conclusions.

Assume, for instance, that the argument game is based on a finite set F of indisputable facts and a finite set R of rules: facts initially fire rules which build proofs for literals. If R and F are common knowledge of Pr and Op, successful strategies in argument games are trivially identified: each player can compute if the entire theory (consisting of F and R) logically entails any l. In this situation the game consists of a single move. The computational complexity of the whole dialogue game reduces to the complexity of deciding the entailment problem in the underlying logic. Suppose now that F is known by both players, but R is partitioned into three subsets: a set R_{Com} known by both players and two subsets R_{Pr} and R_{Op} corresponding, respectively, to Pr's and Op's private knowledge (what Pr and Op privately know to be true). In this context, at each turn a player chooses a theory to prove her claim. A move may use any (i) formulas in R_{Com}, (ii) formulas played in previous turns (by both players), and (iii) selection of formulas (not previously used) from her private knowledge. Hence, we add the complexity of deciding which theory (i.e., subset of her formulas) to play to win the game to the entailment problem. Is there any safe criterion to select successful strategies? Consider a setting where $F = \{a, d, f\}$, $R_{\mathsf{Com}} = \emptyset$, and the players have the following rules:

$$R_{\mathsf{Pr}} = \{a \Rightarrow b, \quad d \Rightarrow c, \quad c \Rightarrow b\}$$
$$R_{\mathsf{Op}} = \{c \Rightarrow e, \quad e, f \Rightarrow \neg b\}.$$

If Pr's intent is to prove b and she plays $\{a \Rightarrow b\}$, then Pr wins the game. If Pr plays $\{d \Rightarrow c,\ c \Rightarrow b\}$ (or even R_{Pr}), this allows Op to succeed. Here, a minimal subset of R_{Pr} is successful. The situation can be reversed for Pr. Replace the sets of private rule with

$$R_{\mathsf{Pr}} = \{a \Rightarrow b, \quad d \Rightarrow \neg c\}$$
$$R_{\mathsf{Op}} = \{d, c \Rightarrow \neg b, \quad f \Rightarrow c\}.$$

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

In the remainder, we will show that the problem of deciding what set of rules to play (Strategic Argumentation Problem) at a given turn is NP-complete even when the problem of deciding whether a given theory (defeasibly) entails a literal can be computed in polynomial time. We will map the NP-complete Restoring Sociality Problem proposed in [7] into the Strategic Argumentation Problem. To this end, we first propose a standard Defeasible Logic to formalise the argumentation framework (Subsection 3.1) and then we present the BIO agent defeasible logic (Subsection 3.2). In Section 5 we show how to

[1] NICTA, Australia
[2] Griffith University, Australia
[3] University of Verona, Italy
[4] CIRSFID University of Bologna, Italy
[5] QUT, Australia

transform an agent defeasible logic into an equivalent argumentation one and we present the main theorem of computational complexity.

2 Related Work

Most existing investigations of strategic argumentation in AI, such as [14, 10, 16, 15, 8] assume that argument games have complete information, which, we noticed, is an oversimplification in many real-life contexts (such as in legal disputes). [10] presents a notion of argument strength within the class of games of strategy while [14, 16, 8] work on two-player extensive-form games of perfect information.

It is still an open problem how to reconstruct formal argumentation with incomplete information in a game-theoretic setting (such as typically in Bayesian extensive games). Preliminary to that, it is crucial to study the computation cost for logically characterizing the problems that any argumentation game with incomplete information potentially rises. Relevant recent papers that studied argumentation of incomplete information without any direct game-theoretic analysis are [11] and [17], which worked within the paradigm of abstract argumentation. The general idea in these works is to devise a system for dynamic argumentation games where agents' knowledge bases can change and where such changes are precisely caused by exchanging arguments. [11] presents a first version of the framework and an algorithm, for which the authors prove a termination result. [17] generalises this framework (by relaxing some constraints) and devises a computational method to decide which arguments are accepted by translating an argumentation framework into logic programming; this further result, however, is possible only when players are eager to give all the arguments, i.e., when proponent and opponent eventually give all possible arguments in the game.

A relevant computational investigation of argumentation-based dialogues is [12]. The underlying formal system of argumentation is again based on Dung's work, but they allow preferences between arguments. The work focuses on persuasion dialogues with incomplete information, which resembles the type of dialogue we model in the present work, as well as the protocol parties use to carry on the argumentation. The analysis shows that given the knowledge base of a party, the problem of determining if a consistent argument to support a given thesis exists is $\mathrm{NP}^{\mathrm{NP}}$-complete. [12] argue that the source of such an high complexity resides in the choice of using standard propositional logic to model the argumentation mechanism. In this respect, they leave an open research question, that is, how exploiting a more efficient mechanism for establishing proofs effectively lower the complexity of the whole problem.

3 Logic

We shall introduce the two logics used in this paper. The first logic, called "Argumentation Logic", reformulates Defeasible Logic (DL) [1] used in a dialogue game to represent the knowledge of the players, the structure of the arguments, and to perform reasoning. [4] provides the relationships between this logic (and some of its variants) and abstract argumentation, and [18] shows how to use this logic for dialogue games. The second logic, called "Agent Logic", was proposed in [7] to model rational agent and to prove that the Restoring Sociality Problem is NP-complete, and is reported in this paper only to express the mechanisms behind our demonstration of NP-completeness.

3.1 Argumentation Logic

Let $\mathrm{PROP}_{\mathrm{arg}}$ be a set of propositional atoms and $\mathrm{Lbl}_{\mathrm{arg}}$ be a set of labels. The set $\mathrm{Lit}_{\mathrm{arg}} = \mathrm{PROP}_{\mathrm{arg}} \cup \{\neg p | p \in \mathrm{PROP}_{\mathrm{arg}}\}$ is the set of *literals*. If q is a literal, $\sim q$ denotes the *complementary* literal: if q is a positive literal p then $\sim q$ is $\neg p$, and if q is $\neg p$, then $\sim q$ is p. A *rule* is an expression of the form $r : \phi_1, \dots, \phi_n \hookrightarrow \psi$, where $r \in \mathrm{Lbl}_{\mathrm{arg}}$ is a unique label, $A(r) = \{\phi_1, \dots, \phi_n\} \subseteq \mathrm{Lit}_{\mathrm{arg}}$ is the set of antecedents of r, $C(r) = \psi \in \mathrm{Lit}_{\mathrm{arg}}$ is the consequent of r, and $\hookrightarrow \in \{\rightarrow, \Rightarrow, \leadsto\}$ is the type of r. $\rightarrow$ denotes *strict rules*, i.e., rules such that whenever the premises are indisputable, so is the conclusion. $\Rightarrow$ denotes *defeasible rules*, i.e., rules that can be defeated by contrary evidence. $\leadsto$ denotes *defeaters*, i.e., rules that are used to prevent some conclusion but cannot be used to draw any conclusion. Given a set of rules R, $R[q]$ indicates all rules in R with consequent q; (i) R_{s}, (ii) R_{d}, (iii) R_{sd}, (iv) and R_{dft} are the subsets of R of (i) strict rules, (ii) defeasible rules, (iii) strict and defeasible rules, (iv) defeaters.

Definition 1. *A* Defeasible Argumentation Theory (DArT) *is a structure $D_{\mathrm{arg}} = (F, R, >)$, where (i) $F \subseteq \mathrm{Lit}_{\mathrm{arg}}$ is a finite set of facts, (ii) R is the finite set of rules, and (iii) $> \subseteq R \times R$ is a binary, acyclic, irreflexive, and asymmetric relation called superiority relation.*

Given a DArT D_{arg}, a proof P of length n in D_{arg} is a finite sequence $P(1), \dots, P(n)$ of *tagged literals* of the type $+\Delta q$, $-\Delta q$, $+\partial q$ and $-\partial q$, with $q \in \mathrm{Lit}_{\mathrm{arg}}$. The proof conditions below define the logical meaning of such tagged literals. $P(1..n)$ denotes the first n steps of proof P.

Given $\# \in \{\Delta, \partial\}$ and a proof P in D_{arg}, a literal q is *#-provable* in D_{arg} at n (or simply *#-provable*) if there is a line $P(m)$ of P such that $m \leq n$ and $P(m) = +\#q$. A literal q is *#-rejected* in D_{arg} at n (or simply *#-rejected*) if there is a line $P(m)$ of P such that $m \leq n$ and $P(m) = -\#q$. We use statements "Δ-provable" (resp. "∂-provable") and "definitely provable" (resp. "defeasibly provable") as synonyms. Similar conventions apply for rejected literals.

In what follows, for space reasons, we only present proof conditions for $+\Delta$ and $+\partial$: the negative ones are obtained via the principle of *strong negation*. This is closely related to the function that simplifies a formula by moving all negations to an inner most position in the resulting formula, and replaces the positive tags with the respective negative tags, and the other way around [2].

The proof conditions for $+\Delta$ describe just forward chaining of strict rules.

$+\Delta$: If $P(n + 1) = +\Delta q$ then
  (1) $q \in F$ or
  (2) $\exists r \in R_{\mathrm{s}}[q]$ s.t. $\forall a \in A(r)$. a is Δ-provable.

Literal q is definitely provable if either (1) is a fact, or (2) there is a strict rule for q, whose antecedents have all been definitely proved.

Definition 2. *Given a proof P in D_{arg}, a rule $r \in R_{sd}$ is (i)* applicable *(at $P(n+1)$) iff $\forall a \in A(r)$, a is #-provable; (ii)* discarded *(at $P(n+1)$) iff $\exists a \in A(r)$ such that a is #-rejected.*

The proof conditions for $+\partial$ are as follows.

$+\partial$: If $P(n + 1) = +\partial q$ then
(1) q is Δ-provable or
(2) (2.1) $\sim q$ is Δ-rejected and
  (2.2) $\exists r \in R_{\mathrm{sd}}[q]$ s.t. r is applicable, and
  (2.3) $\forall s \in R[\sim q]$. either s is discarded, or
    (2.3.1) $\exists t \in R[q]$ s.t. t is applicable and $t > s$.

Literal q is defeasibly provable if (1) q is already definitely provable, or (2) we argue using the defeasible part of the theory. For (2), $\sim q$ is not definitely provable (2.1), and there exists an applicable strict or defeasible rule for q (2.2). Every attack s is either discarded (2.3), or defeated by a stronger rule t (2.3.1).

3.2 Agent Logic

A defeasible agent theory is a standard defeasible theory enriched with (i) modes for rules, (ii) modalities (belief, intention, obligation) for literals, and (iii) relations for conversions and conflict resolution. We report below only the distinctive features, that is, the language and the basics behind the logic. For a detailed exposition see [7].

Let $\mathrm{PROP}_{\mathsf{soc}}$ be a set of propositional atoms, $\mathrm{Lit}_{\mathsf{soc}} = \mathrm{PROP}_{\mathsf{soc}} \cup \{\neg p | p \in \mathrm{PROP}_{\mathsf{soc}}\}$ be the set of literals, $\mathrm{MOD} = \{\mathsf{BEL}, \mathsf{INT}, \mathsf{OBL}\}$ be the set of modal operators, and $\mathrm{Lbl}_{\mathsf{soc}}$ be a set of labels. The set $\mathrm{ModLit} = \{Xl | l \in \mathrm{Lit}_{\mathsf{soc}}, X \in \{\mathsf{OBL}, \mathsf{INT}\}\}$ is the set of *modal literals*. A *rule* is an expression of the form $r : \phi_1, \ldots, \phi_n \hookrightarrow_X \psi$, where $r \in \mathrm{Lbl}_{\mathsf{soc}}$ is a unique label, $A(r) = \{\phi_1, \ldots, \phi_n\} \subseteq \mathrm{Lit}_{\mathsf{soc}} \cup \mathrm{ModLit}$ is the set of antecedents of r, $C(r) = \psi \in \mathrm{Lit}_{\mathsf{soc}}$ is the consequent of r, $\hookrightarrow \in \{\rightarrow, \Rightarrow, \leadsto\}$ is the type of r, $X \in \mathrm{MOD}$ is the mode of r. R^X ($R^X[q]$) denotes all rules of mode X (with consequent q), and $R[q] = \bigcup_{X \in \{\mathsf{BEL}, \mathsf{OBL}, \mathsf{INT}\}} R^X[q]$.

Notice that rules for intention and obligation are meant to introduce modalities: for example, if we have the intention rule $r : a \Rightarrow_{\mathsf{INT}} b$ and we derive a, then we obtain $\mathsf{INT}b$. On the contrary, belief rules produce literals and not modal literals.

We define two relations among different modalities.

Rule conversion. We define an asymmetric binary convert relation $\mathrm{Cv} \subseteq \mathrm{MOD} \times \mathrm{MOD}$ such that $\mathrm{Cv}(Y, X)$ means 'a rule of mode Y can be used also to produce conclusions of mode X'. This corresponds to the following inference rule:

$$\frac{Xa_1, \ldots, Xa_n \quad r\colon a_1, \ldots, a_n \Rightarrow_Y b}{Xb} \ \mathrm{Cv}(Y, X)$$

where $A(r) \neq \emptyset$ and $A(r) \subseteq \mathrm{Lit}$.

Conflict-detection/resolution. We define an asymmetric binary conflict relation $\mathrm{Cf} \subseteq \mathrm{MOD} \times \mathrm{MOD}$ such that $\mathrm{Cf}(Y, X)$ means 'modes Y and X are in conflict and mode Y prevails over X'.

Definition 3. *A* Defeasible Agent Theory (DAgT) *is a structure* $D_{\mathsf{soc}} = (F_{\mathsf{soc}}, R^{\mathsf{BEL}}, R^{\mathsf{INT}}, R^{\mathsf{OBL}}, >_{\mathsf{soc}}, \mathcal{V}, \mathcal{F})$*, where*

- $F_{\mathsf{soc}} \subseteq \mathrm{Lit}_{\mathsf{soc}} \cup \mathrm{ModLit}$ *is a finite set of facts.*
- R^{BEL}, R^{OBL}, R^{INT} *are three finite sets of rules for beliefs, obligations, and intentions.*
- *The superiority (acyclic) relation* $>_{\mathsf{soc}} = >^{sm}_{\mathsf{soc}} \cup >^{\mathrm{Cf}}_{\mathsf{soc}}$ *is such that (i)* $>^{sm}_{\mathsf{soc}} \subseteq R^X \times R^X$ *such that if* $r >_{\mathsf{soc}} s$ *then* $r \in R^X[p]$ *and* $s \in R^X[\sim p]$*; and (ii)* $>^{\mathrm{Cf}}_{\mathsf{soc}}$ *is such that* $\forall r \in R^Y[p], \forall s \in R^X[\sim p]$ *if* $\mathrm{Cf}(Y, X)$ *then* $r >^{\mathrm{Cf}}_{\mathsf{soc}} s$.
- $\mathcal{V} = \{\mathrm{Cv}(\mathsf{BEL}, \mathsf{OBL}), \mathrm{Cv}(\mathsf{BEL}, \mathsf{INT})\}$ *is a set of convert relations.*
- $\mathcal{F} = \{\mathrm{Cf}(\mathsf{BEL}, \mathsf{OBL}), \mathrm{Cf}(\mathsf{BEL}, \mathsf{INT}), \mathrm{Cf}(\mathsf{OBL}, \mathsf{INT})\}$ *is a set of conflict relations.*

A proof is now a finite sequence of tagged literals of the type $+\Delta_X q$, $-\Delta_X q$, $+\partial_X q$ and $-\partial_X q$.

The following definition states the special status of belief rules, and that the introduction of a modal operator corresponds to being able to derive the associated literal using the rules for the modal operator.

Definition 4. *Given* $\# \in \{\Delta, \partial\}$ *and a proof* P *in* D_{soc}*,* q *is* #-provable *(resp.* #-rejected*) in* D *at* n *(or simply* #-provable*, resp.,* #-rejected*) if there is a line* $P(m)$ *of* P *such that* $m \leq n$ *and either*

1. *q is a literal and* $P(m) = +\#_{\mathsf{BEL}} q$ *(resp.* $P(m) = -\#_{\mathsf{BEL}} q$*), or*
2. *q is a modal literal* Xp *and* $P(m) = +\#_X p$ *(resp.* $P(m) = -\#_X p$*), or*
3. *q is a modal literal* $\neg Xp$ *and* $P(m) = -\#_X p$ *(resp.* $P(m) = +\#_X p$*).*

We are now ready to report the proof conditions for $+\Delta_X$.

$+\Delta_X$: If $P(n+1) = +\Delta_X q$ then
(1) $q \in F$ if $X = \mathsf{BEL}$ or $Xq \in F$ or
(2) $\exists r \in R^X_s[q]$ s.t. $\forall a \in A(r)$. a is Δ-provable or
(3) $\exists r \in R^Y_s[q]$ s.t. $\mathrm{Cv}(Y, X) \in \mathcal{V}$ and
$\forall a \in A(r)$. Xa is Δ-provable.

The sole difference with respect to $+\Delta$ is that now we may use rule of a different mode (Y) to derive conclusions of mode X through conversion. In this framework, only belief rules may convert to other modes. Namely the case where every antecedent of the belief rule r in clause (3) must be (definitely) proven with modality X.

We reformulate Definition 2 to take into account Cv and Cf relations.

Definition 5. *Given a proof* P*,* $\# \in \{\Delta, \partial\}$ *and* $X, Y, Z \in \mathrm{MOD}$

- *A rule* r *is* applicable *for* X *(at* $P(n+1)$*) iff*
  1. $r \in R^X$ *and* $\forall a \in A(r)$*,* a *is #-provable, or*
  2. $r \in R^Y$*,* $\mathrm{Cv}(Y, X) \in \mathcal{V}$*, and* $\forall a \in A(r)$*,* Xa *is #-provable.*
- *A rule* r *is* discarded *for* X *(at* $P(n+1)$*) iff*
  3. $r \in R^X$ *and* $\exists a \in A(r)$ *such that* a *is #-rejected; or*
  4. $r \in R^Y$*,* $\mathrm{Cv}(Y, X) \in \mathcal{V}$ *and* $\exists a \in A(r)$ *such that* Xa *is #-rejected, or*
  5. $r \in R^Z$ *and either* $\mathrm{Cv}(Z, X) \notin \mathcal{V}$ *or* $\mathrm{Cf}(Z, X) \notin \mathcal{F}$.

The proof conditions for $+\partial_X$ are the following.

$+\partial_X$: If $P(n+1) = +\partial_X q$ then
(1)Xq is Δ-provable or
(2) (2.1) $X{\sim}q$ is Δ-rejected and
(2.2) $\exists r \in R_{sd}[q]$ s.t. r is applicable, and
(2.3) $\forall s \in R[\sim q]$ either s is discarded, or
(2.3.1) $\exists t \in R[q]$ s.t. t is applicable and $t > s$, and
either $t, s \in R^Z$, or $\mathrm{Cv}(Y, X) \in \mathcal{V}$ and $t \in R^Y$.

Again, the only difference with respect to $+\partial$ is that we have rules for different modes, and thus we have to ensure the appropriate relationships among the rules. Hence, clause (2.3.1) prescribes that either attack rule s and counterattack rule t have the same mode (i.e., $s, t \in R^Z$), or that t can be used to produce a conclusion of the mode X (i.e., $t \in R^Y$ and $\mathrm{Cv}(Y, X) \in \mathcal{V}$). Notice that this last case is reported for the sake of completeness since it plays a role only within theories with more than three modes.

We define the *extension* of a defeasible theory as the set of all positive and negative conclusions. [9, 7] proved that the computing the extension of a theory in both argumentation and agent logic is linear in the size of the theory.

The following notions are needed to formulate the Restoring Sociality Problem [7].

- Given an DAgT D_{soc}, a literal l is *supported* in D_{soc} iff there exists a rule $r \in R[l]$ such that r is applicable, otherwise l is not supported. For $X \in \mathrm{MOD}$ we use $+\Sigma_X l$ and $-\Sigma_X l$ to indicate that l is supported / not supported by rules for X.
- *Primitive intentions* of an agent are those intentions given as facts.
- *Primary* intentions and obligations may not be derived using rule conversion.
- A *social agent* is an agent for which obligation rules are stronger than any conflicting intention rules but weaker than any conflicting belief rules.

3.3 Restoring Sociality Problem

INSTANCE: Let I be a finite set of primitive intentions, $\mathsf{OBL}p$ a primary obligation, and D_{soc} a DAgT modelling a *deviant* agent, i.e. such that $I \subseteq F$, $D_{\mathsf{soc}} \vdash -\partial_{\mathsf{OBL}}p$, $D_{\mathsf{soc}} \vdash -\Sigma_{\mathsf{OBL}}{\sim}p$, $D_{\mathsf{soc}} \vdash +\partial_{\mathsf{INT}}{\sim}p$, $D_{\mathsf{soc}} \vdash +\Sigma_{\mathsf{OBL}}p$ and $D_{\mathsf{soc}} \vdash -\Sigma_{\mathsf{BEL}}{\sim}p$.
QUESTION: Is there a DAgT D'_{soc} equal to D_{soc} but for I' which is a proper subset of I, such that $\forall q$ if $D_{\mathsf{soc}} \vdash +\partial_{\mathsf{OBL}}q$ then $D'_{\mathsf{soc}} \vdash +\partial_{\mathsf{OBL}}q$ and $D'_{\mathsf{soc}} \vdash +\partial_{\mathsf{OBL}}p$?

Let us the consider the DAgT D_{soc} consisting of

$$F = \{\mathsf{INT}p, \mathsf{INT}s\}$$
$$R = \{r_1 : p, s \Rightarrow_{\mathsf{BEL}} q \quad r_2 : \Rightarrow_{\mathsf{OBL}} {\sim}q \quad r_3 : \Rightarrow_{\mathsf{BEL}} s\}$$
$$> = \{r_1 > r_2\}$$

Rule r_1 is a belief rule, which is stronger than the obligation rule r_2 by conflict. In addition, we have that the r_1 is not applicable (i.e., $-\Sigma_{\mathsf{BEL}}q$) since $D_{\mathsf{soc}} \vdash -\partial_{\mathsf{BEL}}p$. There are no obligation rules for q, so $D_{\mathsf{soc}} \vdash -\partial_{\mathsf{OBL}}q$. Rule r_1 behaves as an intention rule since $D_{\mathsf{soc}} \vdash +\partial_{\mathsf{INT}}p$ and $D_{\mathsf{soc}} \vdash +\partial_{\mathsf{INT}}s$. Since r_1 is stronger than r_2, the derivation of $+\partial_{\mathsf{OBL}}{\sim}q$ is prevented against the sociality of the agent.

The related decision problem is whether it is possible to avoid the "deviant" behaviour by giving up some primitive intentions, retaining all the (primary) obligations, and maintaining a set of primitive intentions as close as possible to the original set.

Theorem 6 ([7]). *The Restoring Sociality Problem is NP-complete.*

4 Dialogue Games

The form of a *dialogue game* involves a sequence of interactions between two players, the *Proponent* Pr and the *Opponent* Op. The content of the dispute being that Pr attempts to assess the validity of a particular thesis (called *critical literal*), whereas Op attacks Pr's claims in order to refute such thesis. We point out that in our setting Op has the burden of proof on the opposite thesis, and not just the duty to refute Pr's thesis.

The challenge between the parties is formalised by means of *argument* exchange. In the majority of concrete instances of argumentation frameworks, arguments are defined as chains of reasoning based on facts and rules captured in some formal language (in our case, a defeasible derivation P). Each party adheres to a particular set of game rules as defined below. The players partially share knowledge of a defeasible theory. Each participant has a private knowledge regarding some rules of the theory. Other rules are known by both parties, but this set may be empty. These rules along with all the facts of the theory and the superiority relation represent the common knowledge of both participants. By putting forward a private argument during a step of the game, the agent increases the common knowledge by the rules used within the argument just played.

Define the DArT to be $D_{\mathsf{arg}} = (F, R, >)$ such that (i) $R = R_{\mathsf{Pr}} \cup R_{\mathsf{Op}} \cup R_{\mathsf{Com}}$, (ii) R_{Pr} (R_{Op}) is the private knowledge of Pr (Op), and (iii) R_{Com} is the (possibly empty) set of rules known by both participants. We use the superscript notation D^i_{arg}, R^i_{Pr}, R^i_{Op}, and R^i_{Com} to denote such sets at turn i. D_{arg} is assumed coherent and consistent, i.e., there is no literal p such that: (i) $D_{\mathsf{arg}} \vdash +\partial p$ and $D_{\mathsf{arg}} \vdash -\partial p$, and (ii) $D_{\mathsf{arg}} \vdash +\partial p$ and $D_{\mathsf{arg}} \vdash +\partial{\sim}p$.

We now formalise the game rules which establish how the common theory D^i_{arg} is modified based on the move played at turn i.

The parties start the game by choosing the critical literal l to discuss about: Pr has the burden to prove $+\partial l$ by using the current common knowledge along with a subset of R_{Pr}, whereas Op's final goal is to prove $+\partial{\sim}l$ using R_{Op} instead of R_{Pr}. The players may not present arguments in parallel: they take turn in making their move. The repertoire of moves at each turn just includes 1) putting forward an argument, and 2) passing.

When putting forward an argument at turn i, Pr (Op) may bring a demonstration P whose terminal literal differs from l (${\sim}l$). When a player passes, she declares her defeat and the game ends. This happens when there is no combination of the remaining private rules which proves her thesis.

Hence, the initial state of the game is $D^0_{\mathsf{arg}} = (F, R^0_{\mathsf{Com}}, >)$ with $R^0_{\mathsf{Com}} = R_{\mathsf{Com}}$, and $R^0_{\mathsf{Pr}} = R_{\mathsf{Pr}}$, $R^0_{\mathsf{Op}} = R_{\mathsf{Op}}$. If $D^0_{\mathsf{arg}} \vdash +\partial l$, Op starts the game. Otherwise, the Pr does so.

At turn i, if Pr plays R^i_{arg}, then

- $D^{i-1}_{\mathsf{arg}} \vdash +\partial{\sim}l$ ($D^{i-1}_{\mathsf{arg}} \vdash -\partial l$ if $i = 1$);
- $R^i_{\mathsf{arg}} \subseteq R^{i-1}_{\mathsf{Pr}}$;
- $D^i_{\mathsf{arg}} = (F, R^i_{\mathsf{Com}}, >)$;
- $R^i_{\mathsf{Pr}} = R^{i-1}_{\mathsf{Pr}} \setminus R^i_{\mathsf{arg}}$, $R^i_{\mathsf{Op}} = R^{i-1}_{\mathsf{Op}}$, and $R^i_{\mathsf{Com}} = R^{i-1}_{\mathsf{Com}} \cup R^i_{\mathsf{arg}}$;
- $D^i_{\mathsf{arg}} \vdash +\partial l$.

At turn i, if Op plays R^i_{arg}, then

- $D^{i-1}_{\mathsf{arg}} \vdash +\partial l$;
- $R^i_{\mathsf{arg}} \subseteq R^{i-1}_{\mathsf{Op}}$;
- $D^i_{\mathsf{arg}} = (F, R^i_{\mathsf{Com}}, >)$;
- $R^i_{\mathsf{Pr}} = R^{i-1}_{\mathsf{Pr}}$, $R^i_{\mathsf{Op}} = R^{i-1}_{\mathsf{Op}} \setminus R^i_{\mathsf{arg}}$, and $R^i_{\mathsf{Com}} = R^{i-1}_{\mathsf{Com}} \cup R^i_{\mathsf{arg}}$;
- $D^i_{\mathsf{arg}} \vdash +\partial{\sim}l$.

4.1 Strategic Argumentation Problem

Pr'S INSTANCE FOR TURN i: Let l be the critical literal, R^{i-1}_{Pr} be the set of the private rules of Pr, and D^{i-1}_{arg} be such that either $D^{i-1}_{\mathsf{arg}} \vdash -\partial l$ if $i = 1$, or $D^{i-1}_{\mathsf{arg}} \vdash +\partial{\sim}l$ otherwise.
QUESTION: Is there a subset R^i_{arg} of R^{i-1}_{Pr} such that $D^i_{\mathsf{arg}} \vdash +\partial l$?

Op'S INSTANCE FOR TURN i: Let l be the critical literal, R^{i-1}_{Op} be the set of the private rules of Op, and D^{i-1}_{arg} be such that $D^{i-1}_{\mathsf{arg}} \vdash +\partial l$.
QUESTION: Is there a subset R^i_{arg} of R^{i-1}_{Op} such that $D^i_{\mathsf{arg}} \vdash +\partial{\sim}l$?

5 Reduction

We now prove that the Strategic Argumentation Problem is NP-complete. We start by presenting how to transform a DAgT into a DArT, which requires reframing both literals and rules: whereas the DAgT deals with three different modes of rules and modal literals, the DArT has rules without modes and literals: Definitions 7 and 8 are based on the following ideas:

- To flatten all modal literals with respect to internal negations modalities. For instance, ${\sim}p$ is flattened into the literal *not_p*, while $\mathsf{OBL}q$ is *obl_q*.
- To remove modes from rules for BEL, OBL and INT. Thus, a rule with mode X and consequent p is transformed into a standard, non-modal rule with conclusion Xp. An exception is when we deal with belief rules, given that they do not produce modal literals. Therefore, rule $r : a \Rightarrow_{\mathsf{OBL}} p$ is translated in $r_{fl} : a \Rightarrow obl\_p$, while rule $s : b \Rightarrow_{\mathsf{BEL}} q$ becomes $s_{fl} : b \Rightarrow q$.

Function pflat flattens the propositional part of a literal and syntactically represents negations; function flat flattens modalities.

Definition 7. *Let* D_{soc} *be a DAgT. The transformations* pflat : $\mathrm{Lit}_{\mathsf{soc}} \to \mathrm{PROP}_{\mathsf{arg}}$ *and* flat : $\mathrm{ModLit}_{\mathsf{soc}} \cup \mathrm{Lit}_{\mathsf{soc}} \to \mathrm{Lit}_{\mathsf{arg}}$ *are*

$$\mathrm{pflat}(p) = \begin{cases} p \in \mathrm{PROP}_{\mathsf{arg}} & \textit{if } p \in \mathrm{PROP}_{\mathsf{soc}} \\ not\_q \in \mathrm{PROP}_{\mathsf{arg}} & \textit{if } p = \neg q,\ q \in \mathrm{PROP}_{\mathsf{soc}} \end{cases}$$

$$\mathrm{flat}(p) = \begin{cases} \mathrm{pflat}(q) & \textit{if } p = q, \\ obl\_\mathrm{pflat}(q) & \textit{if } p = \mathsf{OBL}q \\ \neg obl\_\mathrm{pflat}(q) & \textit{if } p = \neg\mathsf{OBL}q \\ int\_\mathrm{pflat}(q) & \textit{if } p = \mathsf{INT}q \\ \neg int\_\mathrm{pflat}(q) & \textit{if } p = \neg\mathsf{INT}q. \end{cases}$$

Given that in the agent logic a belief modal literal is not $\mathsf{BEL}p$ but simply p, we have that $\mathrm{flat}(p) = \mathrm{pflat}(p)$ whenever the considered mode is BEL, while $\mathrm{flat}(Xp) = x\_\mathrm{pflat}(p)$ if $X = \{\mathsf{OBL}, \mathsf{INT}\}$ (and consequently x is obl if $X = \mathsf{OBL}$, int otherwise).

We need to redefine the concept of complement to map modal literals into an argumentation logic with literals obtained through flat. Thus, if $q \in \mathrm{PROP}_{\mathsf{arg}}$ is a literal p then $\sim q$ is not_p; and if q is not_p, then $\sim q$ is p. Moreover, if $q \in \mathrm{Lit}_{\mathsf{arg}}$ is $x\_\mathrm{pflat}(p)$ then $\sim q = x\_\mathrm{pflat}(\sim p)$; and q is $\neg x\_\mathrm{pflat}(p)$ then $\sim q = x\_\mathrm{pflat}(p)$.

We now propose a detailed description of facts and rules introduced by Definition 8. In the Restoring Sociality Problem we have to select a subset of factual intentions, while in the Strategic Argumentation Problem we choose a subset of rules to play to defeat the opponent's argument. Therefore, factual intentions in D_{soc} are modelled in D_{arg} as strict rules with empty antecedent (r_p), while factual beliefs and obligations are facts of D_{arg}.

We recall that, while proving $\pm\#_X q$, a rule in D_{soc} may fire if either is of mode X, through convert, or through conflict. Hence, a rule r in D_{soc} has many counterparts in D_{arg}. Specifically, r_{fl} is built from r by removing the mode and flattening all antecedents, as well as the consequent p which now embeds the mode introduced by r.

Moreover, if $r \in R^{\mathsf{BEL}}[p]$ then it may be used through conversion to derive Xp. To capture this feature, we introduce a rule r_{Cvx} with conclusion $x\_\mathrm{pflat}(p)$ and where for each antecedent $a \in A(r)$ the corresponding in $A(r_{Cvx})$ is $x\_\mathrm{pflat}(a)$ according either to clause (3) of $+\Delta_X$ or to condition 2. of Definition 5.

In D_{soc}, it is easy to determine which rule may fire against one another, being that consequents of rules are non-modal literals. Even when the rules have different modes and the conflict mechanism is used, their conclusions are two complementary literals. Given the definition of complementary literals obtained through flat we have introduced after Definition 7, this is not the case for the literals in D_{arg}. The situation is depicted in the following theory:

$$\begin{array}{ll} r : a \Rightarrow_{\mathsf{OBL}} p & r_{fl} : a \Rightarrow obl\_p \\ s : b \Rightarrow_{\mathsf{INT}} \neg p & s_{fl} : b \Rightarrow int\_not\_p \\ t : c \Rightarrow_{\mathsf{BEL}} p & t_{fl} : c \Rightarrow p. \end{array}$$

Here, r may fire against s through $\mathrm{Cf}(\mathsf{OBL}, \mathsf{INT})$ while r_{fl} cannot, given that obl_p is not the complement of int_not_p. In the same fashion, if we derive $+\partial_{\mathsf{BEL}} c$ then t may fire against s because of $\mathrm{Cf}(\mathsf{BEL}, \mathsf{INT})$, while if we have either $+\partial_{\mathsf{OBL}} c$ or $+\partial_{\mathsf{INT}} c$ then the conflict between beliefs and intentions is activated by the use of r through either $\mathrm{Cv}(\mathsf{BEL}, \mathsf{OBL})$ or $\mathrm{Cv}(\mathsf{BEL}, \mathsf{INT})$, respectively. Nonetheless, in both cases there is no counterpart of t in D_{arg} able to fire against int_not_p.

To obviate this issue, we introduce a defeater r_{CfOI} where (i) we flatten the antecedents of r, and (ii) the conclusion is the intention of the conclusion of r, namely $int\_\mathrm{pflat}(C(r))$. This means that whenever r fires, so does r_{CfOI} attacking s_{fl}. Notice that being r_{CfOI} a defeater, such a rule cannot derive directly $+\partial int\_\mathrm{pflat}(p)$ but just prevents the opposite conclusion. The same idea is adopted for rules r_{Cfbelx} and r_{CvyCfx}: defeaters r_{Cfbelx} are needed to model conflict between beliefs and intentions (as rule t in the previous example), whereas defeaters r_{CvyCfx} take care of situations where $r \in R^Z$ may be used to convert Z into Y and Z prevails over X by Cf.

Thus in the previous example, we would have: (i) $r_{CfOI} : a \rightsquigarrow int\_p$, (ii-iii) $t_{Cfbelx} : c \rightsquigarrow x\_p$, (iv-v) $t_{CvxCfint} : x\_c \rightsquigarrow int\_p$, with $x \in \{obl, int\}$.

Antecedents in D_{soc} may be the negation of modal literals; in that framework, a theory proves $\neg Xp$ if such theory rejects Xp (as stated by condition 3. of Definition 4). In D_{arg} we have to prove $\neg x\_\mathrm{pflat}(p)$ This is mapped in D_{arg} through conditions 8–10 of Definition 8 and the last condition of $>$.

Definition 8. *Let* $D_{\mathsf{soc}} = (F_{\mathsf{soc}}, R^{\mathsf{BEL}}, R^{\mathsf{OBL}}, R^{\mathsf{INT}}, >_{\mathsf{soc}}, \mathcal{V}, \mathcal{F})$ *be a DAgT. Define* $D_{\mathsf{arg}} = (F, R, >)$ *(the* argumentation counterpart *of* D_{soc}*) to be a DArT such that*

$$F = \{\mathrm{flat}(p) | p \in F_{\mathsf{soc}},\ p \in \mathrm{Lit} \textit{ or } p = \mathsf{OBL}q\} \tag{1}$$

$$R = \{r_p :\ \to int\_\mathrm{pflat}(p) | \mathsf{INT}p \in F_{\mathsf{soc}}\} \tag{2}$$

$$\cup \{r_{fl} : \bigcup_{a \in A(r)} \mathrm{flat}(a) \hookrightarrow \mathrm{flat}(p) | r \in R^X[q],\ X = \mathsf{BEL} \textit{ and } p = q, \textit{ or } p = Xq \in \mathrm{ModLit}\} \tag{3}$$

$$\cup \{r_{Cvx} : \bigcup_{a \in A(r)} x\_\mathrm{pflat}(a) \hookrightarrow x\_\mathrm{pflat}(p) | \tag{4}$$

$$r \in R^{\mathsf{BEL}}_{\mathrm{sd}}[p],\ A(r) \neq \emptyset,\ A(r) \subseteq \mathrm{Lit},\ x \in \{obl, int\}\}$$

$$\cup \{r_{CvyCfx} : \bigcup_{y\_\mathrm{pflat}(a) \in A(r_{Cvy})} y\_\mathrm{pflat}(a) \rightsquigarrow x\_\mathrm{pflat}(p) | \tag{5}$$

$$r_{Cvy} \in R[y\_\mathrm{pflat}(p)],\ x, y \in \{obl, int\},\ x \neq y\}$$

$$\cup \{r_{Cfbelx} : \bigcup_{a \in A(r)} \mathrm{flat}(a) \rightsquigarrow x\_\mathrm{pflat}(p) | r \in R^{\mathsf{BEL}}[p],\ x \in \{obl, int\}\} \tag{6}$$

$$\cup \{r_{CfOI} : \bigcup_{a \in A(r)} \mathrm{flat}(a) \rightsquigarrow int\_\mathrm{pflat}(p) | r \in R^{\mathsf{OBL}}[p]\} \tag{7}$$

$$\cup \{r_{-xp} : x\_\mathrm{pflat}(p) \Rightarrow xp | r \in R^Y . \neg Xp \in A(r)\} \tag{8}$$

$$\cup \{r_{-negxp} :\ \Rightarrow \sim xp | r_{-xp} \in R\} \tag{9}$$

$$\cup \{r_{n-xp} : \sim xp \Rightarrow \neg x\_\mathrm{pflat}(p) | r_{-negxp} \in R\} \tag{10}$$

$$> = \{(r_\alpha, s_\beta) | (r, s) \in >_{\mathsf{soc}},\ \alpha, \beta \in \{fl, Cvx, CvxCfy, Cfbelx, CfOI\}\}$$

$$\cup \{(r_{fl}, s_{n-xp}) | r_{fl} \in R[x\_\mathrm{pflat}(p)]\} \tag{11}$$

$$\cup \{(r_{-xp}, s_{-negxp}) | r_{-xp}, s_{dum-negxp} \in R\}.$$

We prove the correctness of the transformation of Definition 8 by showing that such a transformation preserves both positive and negative provability for any given literal.

Theorem 9. *Let* $D_{\mathsf{soc}} = (F_{\mathsf{soc}}, R^{\mathsf{BEL}}, R^{\mathsf{OBL}}, R^{\mathsf{INT}}, >_{\mathsf{soc}}, \mathcal{V}, \mathcal{F})$ *be a DAgT and* D_{arg} *be the argumentation counterpart of* D_{soc}*. Given* $p \in \mathrm{Lit} \cup \mathrm{ModLit}$ *and* $\# = \{\Delta, \partial\}$*:*

1. $D_{\mathsf{soc}} \vdash \pm\#_{\mathsf{BEL}} p$ *iff* $D_{\mathsf{arg}} \vdash \pm\#\mathrm{flat}(p)$*;*
2. $D_{\mathsf{soc}} \vdash \pm\#_X p$ *iff* $D_{\mathsf{arg}} \vdash \pm\#\mathrm{flat}(Xp)$*,* $X \in \{\mathsf{OBL}, \mathsf{INT}\}$.

Proof. For space reasons, see [6] for the full proof. □

To show that the Strategic Argumentation Problem is NP-complete, we have prove that the proposed transformation is polynomial.

Theorem 10. *There is a linear transformation from any DAgT* D_{soc} *to its argumentation counterpart* D_{arg}.

Proof. The transformations of Definition 8 are applied once to each rule and each tuple of the superiority relation. Transformation (1) maps one fact in D_{soc} into one fact in D_{arg}. Transformation (2) maps one primitive intention D_{soc} into one strict rule in D_{arg}. Transformations (3) and (7) again copy one rule into one rule. (4)–(6) generate two rules in D_{arg} for every belief rule in D_{soc}. (8)–(10) generate a total of three rules in D_{arg} for each negative modal literal in D_{soc}. (11) generates thirty-two tuples in D_{arg} for each tuple in $>_{\mathsf{soc}}$ and two tuples for each negative modal literal in D_{soc}.

The above reasoning shows that the transformations perform a number of steps that is, in the worst case, smaller than thirty-two times the size of D_{soc}. This proves the claim. □

Theorem 11. *The Strategic Argumentation Problem is NP-complete.*

Proof. First, the Strategic Argumentation Problem is polynomially solvable on non-deterministic machines since, given a DArT D_{arg}, we guess a set of rules R^i_{arg} and we can check the extension in polynomial time [9]. Second, the Strategic Argumentation Problem is NP-hard. In fact, we map the Restoring Sociality Problem [7] into the Strategic Argumentation Problem. Given a (deviant) DAgT D_{soc}, D_{soc} is mapped into its argumentation counterpart D_{arg} (Definition 8). The transformation is polynomial (Theorem 10) and correct (Theorem 9). □

6 Conclusion

Almost all research in AI on argumentation assumes that strategic dialogues are games of complete information, that is where the structure of the game is common knowledge among the players. Following [11, 17], we argued that argument games work under incomplete information: not knowing the other player's knowledge, each player cannot predict which arguments will be attacked and which counterarguments will be employed for attacking her arguments. We proved that the problem of deciding what set of rules to play at a given move is NP-complete even if the problem of deciding whether a given theory (defeasibly) entails a literal can be computed in polynomial time. To this end, we mapped our problem to the NP-complete Restoring Sociality Problem proposed in [7]. Our research effort is preliminary to a game-theoretic analysis of strategic dialogues, since it studies the computational cost for logically characterising the problem that any argumentation game with incomplete information potentially rises.

In this paper we focused on games with an *asymmetry* with the information shared by the players, but with a *symmetry* on what the two parties have to prove: whereas Pr has to prove l (i.e., $+\partial l$), Op has to prove $\sim l$ (i.e., $+\partial{\sim}l$); however, it is possible to have games where the two parties have a different burden on proof, namely, the proponent Pr has to prove l while the opponent Op has to disprove it. In DL this can be achieved either by proving that the opposite holds, namely $+\partial{\sim}l$ or simply by showing that l is not provable, i.e., $-\partial l$. In this case we have two different types of strategic argumentation problems: one for Pr (which is the same as the current one), and one for Op. For Op, the related decision problem is if there exists a subset of her private rules which, once added to the current set of public rules, allows the resulting theory to prove $-\partial l$. It is easy to understand that such an "attack" is either against l, or against one premise in the derivation. Both share one condition, that is, a rule must change "its status": from being discarded such a rule must become applicable. This is the case only when, given a theory D, its revision D' and an antecedent a, we have that D proves $-\partial a$, while D' proves $+\partial a$. That being said, we argue that the argumentation game to disprove l reduces to the one presented in this paper where the opponent has the burden to prove a. It seems reasonable that even this problem is NP-complete. An investigation of the topic is left for future work.

The NP-completeness result of the paper is proved for the ambiguity blocking, team defeat variant of DL. However, the proof of the result does not depend on the specific features of this particular variant of the logic, and the result extends to the other variants of the logic (see [3] for the definition of the various variants). The version of the argumentation logic presented in this paper does not correspond to the grounded semantics for Dung's style abstract argumentation framework (though it is possible to give such a semantics for it, see [4]). However, the ambiguity blocking variant corresponds to Dung's grounded semantics [4]. Accordingly, strategic argumentation seems to be a computationally infeasible problem in general.

In our game the superiority relation is known *a priori* by both players. If not so, the problem reduces to revising the corresponding Agent Logic by changing a combination of rules and superiority relation. [5] proved that the problem of revising a defeasible theory by only modifying the superiority relation is NP-complete.

REFERENCES

[1] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher, 'Representation results for defeasible logic', *ACM Trans. Comp. Log.*, **2**, 255–287, (2001).

[2] G. Antoniou, D. Billington, G. Governatori, M.J. Maher, and A. Rock, 'A family of defeasible reasoning logics and its implementation', in *ECAI 2000*, pp. 459–463, (2000).

[3] David Billington, Grigoris Antoniou, Guido Governatori, and Michael J. Maher, 'An inclusion theorem for defeasible logics', *ACM Trans. Comput. Log.*, **12**(1), 6, (2010).

[4] G. Governatori, M.J. Maher, G. Antoniou, and D. Billington, 'Argumentation semantics for defeasible logic', *J. Log. Comput.*, **14**(5), 675–702, (2004).

[5] G. Governatori, F. Olivieri, S. Scannapieco, and M. Cristani, 'Revision of defeasible logic preferences', *CoRR*, **abs/1206.5833**, (2012).

[6] G. Governatori, F. Olivieri, S. Scannapieco, A. Rotolo, and M. Cristani, 'Strategic argumentation is NP-complete', *CoRR*, **abs**, (2013).

[7] G. Governatori and A. Rotolo, 'BIO logical agents: Norms, beliefs, intentions in defeasible logic', *Journal of Autonomous Agents and Multi Agent Systems*, **17**(1), 36–69, (2008).

[8] D. Grossi and W. van der Hoek, 'Audience-based uncertainty in abstract argument games', in *IJCAI'13*, pp. 143–149. AAAI Press, (2013).

[9] M.J. Maher, 'Propositional defeasible logic has linear complexity', *TPLP*, **1**(6), 691–711, (2001).

[10] P. Matt and F. Toni, 'A game-theoretic measure of argument strength for abstract argumentation', in *JELIA 2008*, volume 5293 of *LNCS*, pp. 285–297. Springer, (2008).

[11] K. Okuno and K. Takahashi, 'Argumentation system with changes of an agent's knowledge base', in *IJCAI'09*, pp. 226–232, (2009).

[12] Simon Parsons, Michael Wooldridge, and Leila Amgoud, 'Properties and complexity of some formal inter-agent dialogues', *J. Log. Comput.*, **13**(3), 347–376, (2003).

[13] Henry Prakken, 'An abstract framework for argumentation with structured arguments', *Argument & Computation*, **1**(2), 93–124, (2010).

[14] A.D. Procaccia and J.S. Rosenschein, 'Extensive-form argumentation games.', in *EUMAS 2005*, pp. 312–322. Koninklijke Vlaamse Academie van Belie voor Wetenschappen en Kunsten, (2005).

[15] I. Rahwan and K. Larson, 'Argumentation and game theory', in *Argumentation in Artificial Intelligence*. Springer, (2009).

[16] R. Riveret, H. Prakken, A. Rotolo, and G. Sartor, 'Heuristics in argumentation: A game theory investigation', in *COMMA 2008*, pp. 324–335. IOS Press, (2008).

[17] K. Satoh and K. Takahashi, 'A semantics of argumentation under incomplete information', in *Proceedings of Jurisn 2011*, (2011).

[18] S. Thakur, G. Governatori, V. Padmanabhan, and J. Eriksson Lundström, 'Dialogue games in defeasible logic', in *Australian Conference on Artificial Intelligence*, pp. 497–506, (2007).

[19] F. Toni, 'A generalised framework for dispute derivations in assumption-based argumentation', *Artif. Intell.*, **195**, 1–43, (2013).