Strategic Argumentation

Guido Governatori
Data61-CSIRO, Brisbane, Australia
guido.governatori@data61.csiro.au

Michael J. Maher
Reasoning Research Institute, Canberra, Australia
michael.maher@reasoning.org.au

Francesco Olivieri
Griffith University, Brisbane, Australia
francesco.olivieriphd@gmail.com

Abstract

Dialogue games are a dynamic form of argumentation, with multiple parties pooling their arguments with the intention of settling an issue. Such games can have a variety of structures, and may be collaborative or competitive, depending on the motivations of the parties. Strategic argumentation is a class of competitive dialogue games in which two players take turns in contributing their arguments, each attempting to have an issue settled in the way that they would prefer. Thus strategic argumentation games are less about discovering a joint truth than about a player imposing their view on an opponent. They are reflective of legal argumentation.

In the games we study, players have perfect information of the moves players make, but incomplete information on the possible moves (arguments) that other players have available to them. We look both at games using logically structured arguments and games using abstract arguments. We show that playing these games can be computationally hard. We also examine issues of corruption in such games, and discuss approaches to foiling it.

1 Introduction

When two or more parties are engaged in a debate, it is often the case that each party has some information they are not willing to disclose to the other parties. Also,

in some cases, the disclosure of some piece of information by one party could prove detrimental for the party, in the sense that the information could be used to prevent the party to reach their aim in the debate, or some of the information disclosed can help the other party to achieve their goal. Accordingly we can provide the following (informal) definition of strategic argumentation.

Definition 1.1. Strategic argumentation *is the problem of determining what arguments (pieces of information) to disclose during a debate in order to achieve the aim a party has in the debate and to prevent the other party from gaining an undesired advantage.*

To illustrate the issue, consider the following argument exchange, first proposed in [124]:

Example 1.2. *Let* Pr *and* Op *be the players involved in the following argumentation dialogue (*Pr *and* Op *denote, respectively, the proponent and the opponent):*

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 is 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 did not go there. I was at facility A at that time."*
Pr_3 : *"At facility A? Then, it is impossible to have had your ID card stolen since facility A does not allow a person to enter without an ID card."*

We can easily represent arguments of this example with a rule-based formalism as follows. We have rules R:

$$
\begin{array}{rrcl}
r_{\mathsf{Pr}_0}: & & \Rightarrow & murderer(X) \\
r'_{\mathsf{Op}_1}: & & \Rightarrow & \neg evidence\_Against(X) \\
r''_{\mathsf{Op}_1}: & \neg evidence\_Against(X) & \Rightarrow & \neg murderer(X) \\
r_{\mathsf{Pr}_1}: & ID(X)\_at\_crime\_scene & \Rightarrow & evidence\_Against(X) \\
r_{\mathsf{Op}_2}: & ID(X)\_stolen & \Rightarrow & \neg evidence\_Against(X) \\
r'_{\mathsf{Pr}_2}: & & \Rightarrow & only(X)\_at\_crime\_scene \\
r''_{\mathsf{Pr}_2}: & only(X)\_at\_crime\_scene & \Rightarrow & murderer(X) \\
r_{\mathsf{Op}_3}: & at\_facility\_A(X) & \Rightarrow & \neg only(X)\_at\_crime\_scene \\
r_{\mathsf{Pr}_3}: & at\_facility\_A(X) & \Rightarrow & \neg ID(X)\_stolen
\end{array}
$$

and a priority relation $>= \{r_{\mathsf{Op}_2} > r_{\mathsf{Pr}_1}\}$, *where the notation* $r_i : A(r) \Rightarrow C(r)$ *identifies that* r_i *is the name of the rule,* $A(r)$ *is the set of antecedents (possibly empty)*

while $C(r)$ is the conclusion, symbol $\Rightarrow$ denotes that the conclusion may be defeated by contrary evidence, as for instance the conflict between r_{Op_2} and r_{Pr_1}, resolved by $>$ (the superiority relation) which allows us to conclude that $\neg evidence\_Against(X)$ is the case.

A feature of this dialogue is that the exchange of arguments reflects an asymmetry of information between the two parties. Each player does not know the other player's knowledge, thus they cannot predict which arguments will be attacked, nor which counterarguments may be employed for attacking their own arguments. In addition, the private information disclosed by a party might eventually be used by the adversary to construct and play justified counterarguments. Thus, Pr_3 attacks Op_2, but only after Op_3 has been given. Thus, the attack Pr_3 of the proponent is possible *only when* the opponent discloses some private information through the move Op_3 (in this setting, after Op let Pr know that Op was at facility). If we assume that Pr wishes to expose Op's guilt, and Op wishes to hide it, then we can view this dialogue as a game, where a move consists of stating an argument.

This example illustrates a scenario where some of the information disclosed by a party could be detrimental to their aim. This is a common phenomenon in many applications that are suitable to be formally represented by argumentation such as negotiation [117], brokering [10], and in the legal domain [114; 63]. In a negotiation, the other party could use the information to gain some advantage either on the issue of the negotiation (e.g., price of an item) or on some side effects; in a legal proceeding the opposite party could use the information to win the case. Hence, players in such an argumentation game must be strategic in what arguments they expose, to put themselves in the best position. We refer to such games as *strategic argumentation* games.

Furthermore, in such applications the parties can be represented by agents acting and debating on behalf of their clients, but these agents might not have their client's best interests at heart. This can corrupt the dialogue. For example, suppose the agent for Pr was bribed by Op to omit the claim Pr_2. Then Op_3 would have remained private, and Op's lie would be undiscovered. Similar issues occur whenever we employ an agent, whether human or software.

Technically, games involving privacy are called games of *incomplete information*. As argued in [67], argument games with incomplete information can be modelled by stating that each player has a logical theory, constituting their private knowledge, and which is unknown by the opposite party, and there is an additional theory shared by all parties with the information that is public. A player may build an argument that supports their claim by using some of their private knowledge and the common information; in turn, the other party may construct new arguments by re-using

the adversary's disclosed information (along with other pieces of their own private knowledge) in order to defeat the opponent's arguments. In a legal proceeding, we can distinguish between two types of information: the norms in force in the underlying jurisdiction, which are assumed to be known by both parties, and the information, private to each party, on the facts of the case. Accordingly, the legal proceeding can be modelled by three theories, a public one with the common knowledge, encoding the norms of the underlying jurisdiction, plus two private theories: one for each party.

When working with logically structured arguments, the different logical theories are represented by sets of rules (which may include unconditional facts). So, the set R of all rules used to build arguments is partitioned into three (distinct) 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. While the game is evolving, at each turn, a party discloses some of their private arguments and, by doing so, the player reduces their private information ($R_{\mathsf{Pr}}/R_{\mathsf{Op}}$ decreases) with what now becomes part of the new common knowledge base (R_{Com} increases). Consider a setting where $F = \{a, d, f\}$ is the known set of facts (categorical statements), $R_{\mathsf{Com}} = F$ (the facts are common knowledge), and the players have the following sets of rules:

$$R_{\mathsf{Pr}} = \{r_0 : a \Rightarrow b;\ r_1 : d \Rightarrow c;\ r_2 : c \Rightarrow b\} \quad R_{\mathsf{Op}} = \{r_3 : c \Rightarrow e;\ r_4 : e, f \Rightarrow \neg b\}.$$

If Pr's intent is to prove b and plays $\{a \Rightarrow b\}$, then Pr wins the game. In fact, Op has no way to prove e and thus r_4 is not active. If, on the other hand, Pr plays $\{d \Rightarrow c,\ c \Rightarrow b\}$ (or even the whole 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 rules with

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

Playing $\{a \Rightarrow b\}$ is now not successful for Pr, while the whole R_{Pr} ensures victory.

Example 1.2 brings out the issues we will address in this chapter: formalizing such dialogues as strategic argumentation games, addressing the difficulty of making a move in a game, and examining the possibility of corruption in such games and means to foil it. We will look at both defeasible logics [6] and ASPIC-style structured argumentation [2; 111] as languages for expressing arguments. We will also show that the same issues arise if we formulate strategic argumentation in terms of abstract arguments [41]. In looking at corruption, we consider two forms: *espionage* and *collusion*. To counter these possibilities, we examine the use of standards and audit to limit the ability of players to behave corruptly, and the idea of computational resistance to corruption to discourage corruption.

The layout of this chapter is as follows. Section 2 describes the general setting of argumentation and dialogue games. Section 3 provides some technical background on computational complexity, elements of abstract argumentation [41], and a framework for argumentation with logically structured arguments. Section 4 outlines Defeasible Logic and its four main variants. Section 5 presents an instance of the strategic argumentation game with Defeasible Logic as the basis for argumentation, and proves the computational difficulty of playing the game. It extends this result to an instance of structured argumentation under the grounded semantics. Section 6 extends the idea of strategic argumentation further, to abstract argumentation over a variety of semantics. Section 7 investigates how corruption can affect argumentation games, and how it can be countered. Section 8 discusses related work and Section 9 considers possible future directions of this research. Section 10 ends the chapter.

2 Argumentation and Dialogue Games

In this section we briefly describe a general setting of argumentation and dialogue games. In doing so we will not bind concepts such as *argument*, *aim*, *acceptance* or *extension* to a specific meaning, nor specify all details of concepts like argumentation framework. They will be specified more precisely later.

Definition 2.1 (Argumentation framework)**.** *An* argumentation framework *AF is a tuple $(\mathcal{A}, \mathcal{R})$, where $\mathcal{A}$ is a set of arguments, and $\mathcal{R}$ is a collection of relations over $\mathcal{A}$.*

The literature in argumentation theory flourishes with different frameworks describing *what* arguments are, where the two main school of thoughts see them as either *monads* (with no internal structure), or *structured* (made of sub-parts). We will address both schools. For now, we are only interested in saying that there is a function mapping arguments to elements of the language, referred to as *conclusions* (or theses, claims).

Definition 2.2 (Conclusions)**.** *Given an argumentation framework AF and a language of expressions L, the function $conc : \mathcal{A} \mapsto 2^L$ maps each argument to a set of elements of L. If $c_A \in conc(A)$, then we call c_A a* conclusion *of argument A.*

In the monadic view, each argument might have a single, distinct conclusion. In that case, conclusions add nothing to the argumentation framework. In the structured view, an expression might be a conclusion of several arguments, and its negation might also be a conclusion of arguments. Any structured argumentation framework with conclusions can be abstracted to a monadic argumentation framework by simply ignoring its internal structure (and retaining the conclusion function).

For the purposes of this chapter, a semantics maps an argumentation framework to a set of extensions. Each semantics implicitly expresses a criterion for how arguments can coherently be adjudicated together, given an argumentation framework. Each extension in the semantics represents a "reasonable" adjudication, according to that criterion, of the arguments in the argumentation framework. We leave open the details of what an extension is and how it might be represented, but commonly it is a set of arguments or a labelling for arguments (see Section 3.2 for more details of these common representations).

Definition 2.3 (Semantics)**.** *A* semantics *is a function σ mapping argumentation frameworks to a set of extensions.*

There is a rich array of interactions that are considered dialogues in the argumentation literature [24] but, as can be seen from the introduction, we have a specific kind of dialogue in mind. We define a *dialogue* as the exchange of arguments between two (or more) parties. We talk of *dialogue games* when we want to analyse the formal properties of the dialogue, using criteria from game theory.

At the beginning of a dialogue game, each agent starts with a private set of arguments but they also share a (possibly empty) set of arguments that are common knowledge[1] to all players. This shared knowledge among the agents will be enriched throughout the game with the arguments played at each turn, as will be clear in the following.

Each player also has an *aim*, the details of which we leave open. Aims might be to have a particular argument accepted in at least one extension, under a particular semantics, or to have the cardinality of each extension, under a given semantics, be a prime number[2].

Our dialogue games consist of a state and possible changes of state.

Definition 2.4 (Dialogue Game State)**.** *Given a set of agents $Pl_1, \ldots, Pl_n$ (referred to as* players*), a* dialogue game state *is an argumentation framework $(\mathcal{A}, \mathcal{R})$ where $\mathcal{R}$ contains unary relations $\xi_1, \ldots \xi_n$ on $\mathcal{A}$, one for each player, as well as ξ_{Com} and, possibly, other relations.*

Each unary relation ξ_i defines a subset S_i of $\mathcal{A}$: $S_i = \{a \mid a \in \mathcal{A}, \xi_i(a)\}$. Similarly, $S_{\mathsf{Com}} = \{a \mid a \in \mathcal{A}, \xi_{\mathsf{Com}}(a)\}$. S_{Com} is the set of arguments that are common knowledge to all players, while S_i is the additional set of arguments that player Pl_i knows, but other players don't know she knows (they are *private*).

[1] By common knowledge we mean, not only that all players have knowledge of the arguments, but also each player knows that the others know, and each knows that the others know that she knows, and so on. [49]

[2] Admittedly, the latter example is not likely to arise in practice.

Thus, a dialogue game state can equally be viewed as a *split argumentation framework* $(\mathcal{A}, S_{\mathsf{Com}}, S_1, \ldots, S_n, \mathcal{R}')$, where $S_{\mathsf{Com}} \cap (\cup_{i=1}^{n} S_i) = \emptyset$ and $\mathcal{R}'$ is $\mathcal{R} \backslash \{\xi_{\mathsf{Com}}, \xi_1, \ldots, \xi_n\}$.

A dialogue game is a collection of players, each with their own aim, making moves, in turn, to achieve their aim[3].

Definition 2.5 (Dialogue Game)**.** *Given a set of players $Pl_1, \ldots, Pl_n$ and an aim for each player, a* dialogue game *consists of an initial dialogue game state in the form of a split argumentation framework $(\mathcal{A}, S_{\mathsf{Com}}, S_1, \ldots, S_n, \mathcal{R})$, and state transition rules (or* moves*) defined below.*

1. *Players take turns, meaning that only a single player can act at a given turn*[4]*.*
2. *At a given turn k, player Pl_i advances a subset T of its private arguments in order to achieve their aim. If S_{Com}^{k-1} and S_i^{k-1} denote, respectively, the common shared argumentation framework and Pl_i's private argumentation framework at turn $k-1$, then*
   - $S_{\mathsf{Com}}^{k} = S_{\mathsf{Com}}^{k-1} \cup T$
   - $S_i^k = S_i^{k-1} \setminus T$
   - $S_j^k = S^{k-1}$ *for* $j \neq i$
3. *The game ends at turn $k+1$, when either: (i) the aim of each player is satisfied, so no player has an incentive to change the state of the game, or (ii) no player with an unsatisfied aim is able to satisfy that aim by making a move.*

The state of the dialogue game after turn k is $(\mathcal{A}, S_{\mathsf{Com}}^k, S_1^k, \ldots, S_n^k, \mathcal{R})$. The common argumentation framework at that point is $CAF^k = (S_{\mathsf{Com}}^k, \mathcal{R})$.

According to the typology of argumentation games in [128], these dialogue games have a dialectical argumentation mechanism and players have no awareness of other players' arguments; agent type is not specified. The games we define below (Definitions 2.6 and 2.7) have an indicator agent type.

[3]Many different types of dialogue have been classified and many protocols have been provided for them; we refer to Chapter 9 of the present volume [24] for in depth analysis of the various alternatives. In this chapter we restrict ourselves to a minimal and limited view of dialogue games, suitable to define strategic argumentation.

[4]We shall not dwell on the details of how/which players are selected to act at a given turn, as it is outside the scope of this chapter. [128] discusses some other possibilities.

If we ignore turn-taking, our dialogue games are *memoryless*: the permitted moves are determined by the current dialogue state, independent of how that state was reached. Other forms of dialogue game may not have this property.

Note that, although the set of common arguments increases monotonically, this game is non-monotonic, meaning that, at any given turn, aims that were satisfied at the previous turn might now be unsatisfied.

Also note that we are considering the relations $\mathcal{R}$ to have a fixed meaning, independent of player's beliefs or perceptions. The *omniscient* argumentation framework corresponding to a dialogue game is $(\mathcal{A}, \mathcal{R})$.

We now formulate a specific type of dialogue games, namely *strategic argumentation dialogues*. In a strategic argumentation dialogue game, we have only *two* players, who take turns in exchanging arguments to accept/reject a topic φ, where $\varphi \in L$. We name one player *Proponent* (Pr), and the other *Opponent* (Op). We shall consider two variants of the strategic argumentation dialogue game: the *symmetric*, and the *asymmetric* strategic argumentation dialogue game. In the symmetric variant, both parties have the burden of proof, that is, the proponent has to establish φ, where the opponent has to establish $\neg\varphi$. (With $\neg\varphi$, we denote the contrary of φ.) In the asymmetric variant, the proponent still has to establish φ, whereas the opponent aims to prevent this.

In the symmetric variant, at one turn, either φ, or $\neg\varphi$, is accepted. If φ is accepted, then it is the opponent's turn; if $\neg\varphi$ is accepted, then is the proponent's turn. At a given turn, the player has two possible courses of action. First, they play a subset of their private argumentation framework (i.e., a non-empty set of arguments). By doing so, they increment the shared argumentation framework with the arguments just played. Second, they pass and admit defeat. This happens when they are not able to change the status of the conclusion. The game ends when a player passes.

Definition 2.6 (Symmetric Strategic Argumentation Dialogue Game)**.** *Consider two players, a proponent* Pr *and an opponent* Op*, an expression* $\varphi \in L$*, a dialogue game state in the form of a split argumentation framework* $(\mathcal{A}, S_{\mathsf{Com}}, S_{\mathsf{Pr}}, S_{\mathsf{Op}}, \mathcal{R})$*, and a conclusion function conc. Suppose that there is an argument* $a \in S_{\mathsf{Pr}}$ *such that* $\varphi \in conc(a)$*.*

Let S^k_{Com}*,* S^k_{Pr}*, and* S^k_{Op} *denote, respectively, the common knowledge arguments,* Pr*'s private arguments and* Op*'s private arguments after turn* k*. (In particular,* $S^0_{\mathsf{Com}} = S_{\mathsf{Com}}$*,* $S^0_{\mathsf{Pr}} = S_{\mathsf{Pr}}$*, and* $S^0_{\mathsf{Op}} = S_{\mathsf{Op}}$*.)*

We define a symmetric strategic argumentation dialogue game *as a dialogue game where:*

1. *The players take turns; if* φ *is accepted by* CAF^0 *under semantics* σ*, then* Op

begins; otherwise Pr *does so.*

2. *At turn k, if $\neg\varphi$ is accepted in CAF^{k-1}, then it is* Pr*'s turn to play, as follows*

   - Pr *advances a subset of its private arguments $T \subseteq S^{k-1}_{\mathsf{Pr}}$ so that φ is accepted in CAF^k. As a result*
     - $S^k_{\mathsf{Com}} = S^{k-1}_{\mathsf{Com}} \cup T$;
     - $S^k_{\mathsf{Pr}} = S^{k-1}_{\mathsf{Pr}} \setminus T$.
     - $S^k_{\mathsf{Op}} = S^{k-1}_{\mathsf{Op}}$

3. *At turn k, if φ is accepted in CAF^{k-1}, then it is* Op*'s turn to play, as follows*

   - Op *advances a subset of its private arguments $T \subseteq S^{k-1}_{\mathsf{Op}}$ so that $\neg\varphi$ is accepted in CAF^k. As a result*
     - $S^k_{\mathsf{Com}} = S^{k-1}_{\mathsf{Com}} \cup T$;
     - $S^k_{\mathsf{Pr}} = S^{k-1}_{\mathsf{Pr}}$
     - $S^k_{\mathsf{Op}} = S^{k-1}_{\mathsf{Op}} \setminus T$.

4. *The game ends at turn $k+1$, when either (i) it is* Pr*'s turn and there is no move for* Pr *such that CAF^{k+1} accepts φ, in which case* Op *wins, or (ii) it is* Op*'s turn and there is no move for* Op *such that CAF^{k+1} accepts $\neg\varphi$, in which case* Pr *wins.*

The only difference in the asymmetric variant with respect to the symmetric one is that, the opponent no longer has the burden of proof: during her turn, Op proposes arguments in order to prevent acceptance of φ, rather than to accept $\neg\varphi$ (see point 3).

Definition 2.7 (Asymmetric Strategic Argumentation Dialogue Game)**.** *Consider two players, a proponent* Pr *and an opponent* Op*, an expression $\varphi \in L$, a dialogue game state in the form of a split argumentation framework $(\mathcal{A}, S_{\mathsf{Com}}, S_{\mathsf{Pr}}, S_{\mathsf{Op}}, \mathcal{R})$, and a conclusion function conc.*

Let S^k_{Com}, S^k_{Pr}, and S^k_{Op} denote, respectively, the common knowledge arguments, Pr*'s private arguments and* Op*'s private arguments after turn k. (In particular, $S^0_{\mathsf{Com}} = S_{\mathsf{Com}}$, $S^0_{\mathsf{Pr}} = S_{\mathsf{Pr}}$, and $S^0_{\mathsf{Op}} = S_{\mathsf{Op}}$.)*

We define an asymmetric strategic argumentation dialogue game *as a dialogue game where:*

1. *The players take turns; if φ is accepted by CAF^0 under semantics σ, then* Op *begins; otherwise* Pr *does so.*

2. *At turn k, if φ is not accepted in CAF^{k-1}, then it is* Pr*'s turn to play, as follows*

   - Pr *advances a subset of its private arguments $T \subseteq S^{k-1}_{\mathsf{Pr}}$ so that φ is accepted in CAF^k. As a result*
     - $S^k_{\mathsf{Com}} = S^{k-1}_{\mathsf{Com}} \cup T$;
     - $S^k_{\mathsf{Pr}} = S^{k-1}_{\mathsf{Pr}} \setminus T$.
     - $S^k_{\mathsf{Op}} = S^{k-1}_{\mathsf{Op}}$

3. *At turn k, if φ is accepted in CAF^{k-1}, then it is* Op*'s turn to play, as follows*

   - Op *advances a subset of its private arguments $T \subseteq S^{k-1}_{\mathsf{Op}}$ so that φ is not accepted in CAF^k. As a result*
     - $S^k_{\mathsf{Com}} = S^{k-1}_{\mathsf{Com}} \cup T$;
     - $S^k_{\mathsf{Pr}} = S^{k-1}_{\mathsf{Pr}}$
     - $S^k_{\mathsf{Op}} = S^{k-1}_{\mathsf{Op}} \setminus T$.

4. *The game ends at turn $k+1$, when either (i) it is* Pr*'s turn and there is no move for* Pr *such that CAF^{k+1} accepts φ, in which case* Op *wins, or (ii) it is* Op*'s turn and there is no move for* Op *such that CAF^{k+1} does not accept φ, in which case* Pr *wins.*

Thus both variants are dialogue games between two players arguing about a conclusion φ on the basis of their common argumentation framework. They leave open the notion of acceptance and the details of the set of relations $\mathcal{R}$, but specify more precisely the aims of the players. From now on, we will use the abbreviations SSA for Symmetric Strategic Argumentation, and AsSA for Asymmetric Strategic Argumentation.

The asymmetric game can be seen in situations where the parties have different proof standards. For example, in a criminal proceeding the prosecution must prove its case "beyond a reasonable doubt", while the defence has only to prevent this. An asymmetric dialogue game was presented in [48].

The problems that the players must solve in order to move vary slightly according to the kind of game played (SSA vs. AsSA) and the players (Pr and Op). We formulate them as decision problems as follows:

SSA Problem under Semantics σ

Let $(\mathcal{A}, S^k_{\mathsf{Com}}, S^k_{\mathsf{Pr}}, S^k_{\mathsf{Op}}, \mathcal{R})$ be the split argumentation framework as in Definition 2.6 after turn k, and $\varphi \in L$ be the content of the dispute.

Pr's instance for turn $k+1$: A split argumentation framework $(\mathcal{A}, S^k_{\mathsf{Com}}, S^k_{\mathsf{Pr}}, S^k_{\mathsf{Op}}, \mathcal{R})$ and an expression $\varphi \in L$.

Question: Does there exist a subset T of S^k_{Pr} such that φ is accepted by CAF^{k+1} under semantics σ?

Op's instance for turn $k+1$: A split argumentation framework $(\mathcal{A}, S^k_{\mathsf{Com}}, S^k_{\mathsf{Pr}}, S^k_{\mathsf{Op}}, \mathcal{R})$ and an expression $\varphi \in L$.

Question: Does there exist a subset T of S^k_{Op} such that $\neg\varphi$ is accepted by CAF^{k+1} under semantics σ?

Analogously, we can formalise the AsSA Problem.

AsSA Problem under Semantics σ

Let $(\mathcal{A}, S^k_{\mathsf{Com}}, S^k_{\mathsf{Pr}}, S^k_{\mathsf{Op}}, \mathcal{R})$ be the split argumentation framework as in Definition 2.7 after turn k, and $\varphi \in L$ be the content of the dispute.

Pr's instance for turn $k+1$: A split argumentation framework $(\mathcal{A}, S^k_{\mathsf{Com}}, S^k_{\mathsf{Pr}}, S^k_{\mathsf{Op}}, \mathcal{R})$ and an expression $\varphi \in L$.

Question: Does there exist a subset T of S^k_{Pr} such that φ is accepted by CAF^{k+1} under semantics σ?

Op's instance for turn $k+1$: A split argumentation framework $(\mathcal{A}, S^k_{\mathsf{Com}}, S^k_{\mathsf{Pr}}, S^k_{\mathsf{Op}}, \mathcal{R})$ and an expression $\varphi \in L$.

Question: Does there exist a subset T of S^k_{Op} such that φ is not accepted by CAF^{k+1} under semantics σ?

In Section 5, we will give an implementation of the strategic argumentation game with *Defeasible Logic* (DL) [104] as the underlying logical framework, and assess the complexity of these problems.

3 Background

In this section we outline the concepts we use involving computational complexity, abstract and structured argumentation. This is not intended to be an introduction to these topics, it is simply a sketch of the concepts, assuming a familiarity with the more common elements. Those with less familiarity with these topics might want to read an introduction first, such as [75; 45] for computational complexity, [13; 12] for abstract argumentation, and [112] for structured argumentation.

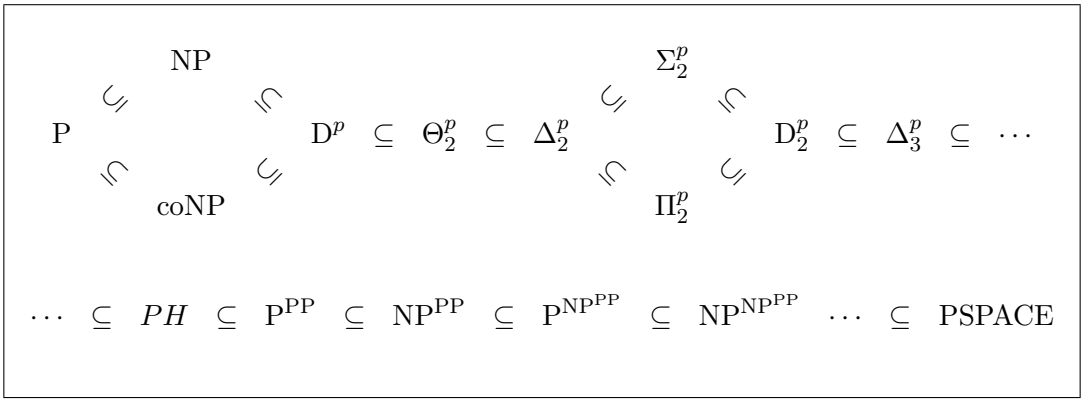

Figure 1: Some complexity classes in the polynomial counting hierarchy, ordered by containment.

3.1 Complexity Classes

When addressing computational complexity we will focus on decision problems, because of their more familiar complexity classes, rather than their functional counterparts, which are more appropriate for many of the computational tasks we will address. We assume familiarity with the polynomial time complexity hierarchy but we will introduce some other complexity classes that we will need, and computational problems that are complete for each class. As is usual, $\mathcal{D}^{\mathcal{C}}$ denotes the class of problems that can be solved with complexity $\mathcal{D}$ if given an oracle for a problem in $\mathcal{C}$.

Within the polynomial hierarchy, a complete problem for Σ_n^p (Π_n^p) is the satisfiability of quantified Boolean formulas (QBF) with quantifiers in the form $\exists\forall\exists\cdots\exists$ (respectively, $\forall\exists\forall\cdots\exists$) with n alternations of quantifiers. PSPACE is the class of decision problems solvable in polynomial space. It contains the entire polynomial hierarchy PH. A complete problem for PSPACE is satisfiability of all quantified Boolean formulas.

D^p is the complexity class of problems that can be expressed as the conjunction of a problem in NP and a problem in coNP. A complete problem for D^p asks, given Boolean formulas ϕ and ψ, is ϕ unsatisfiable and ψ satisfiable? $NP^{D^p} = \Sigma_2^p$. Similarly D_2^p is the conjunction of problems in Σ_2^p and Π_2^p.

Θ_2^p is the class of decision problems solvable by a deterministic polynomial algorithm with $O(log\, n)$ calls to an NP oracle. It is equal to $P_{||}^{NP}$, the class of problems solvable by a deterministic polynomial algorithm with non-adaptive calls to an NP oracle. Non-adaptive refers to the restriction that oracle calls cannot depend on the outcome of previous calls. $NP^{\Theta_2^p} = \Sigma_2^p$.

Δ_2^p is equal to P^{NP}. A complete problem for Δ_2^p is, given a Boolean formula ψ,

does the lexicographically last satisfying assignment for ψ end with a 1?

PP is, roughly, the class of decision problems that have more accepting paths than rejecting paths. It can be thought of as a decision problem version of the more familiar counting complexity class #P, which addresses absolute counting, while PP addresses relative size of counts. We have $\mathrm{P}^{\#\mathrm{P}} = \mathrm{P}^{\mathrm{PP}}$ and $\mathrm{NP}^{\#\mathrm{P}} = \mathrm{NP}^{\mathrm{PP}}$. The entire polynomial hierarchy is contained within $\mathrm{NP}^{\mathrm{PP}}$. A complete problem for PP, called MAJSAT, is to decide whether a given Boolean formula is satisfied by more than half of the assignments to its variables. This can be expressed via a "counting" quantifier C as satisfying $\mathsf{C}X\ \psi$. Similarly, a complete problem for $\mathrm{NP}^{\mathrm{PP}}$, called E-MAJSAT is satisfying formulas $\exists X \mathsf{C} Y\ \psi$. And so on.

The *counting polynomial hierarchy* [137] extends the polynomial hierarchy by incorporating PP, P^{PP}, $\mathrm{NP}^{\mathrm{PP}}$, $\mathrm{coNP}^{\mathrm{PP}}$, etc. Figure 1 displays containment relations among relevant complexity classes. In addition to the containments displayed, $\Theta_2^p \subseteq \mathrm{PP} \subseteq \mathrm{P}^{\mathrm{PP}}$.

3.2 Abstract Argumentation

Definition 3.1 (Abstract Argumentation Framework)**.** *An abstract argumentation framework is a pair* $(\mathcal{A}, \gg)$ *where* $\mathcal{A}$ *is a set of arguments and* $\gg$ *is a subset of* $\mathcal{A} \times \mathcal{A}$*, where* $(a, b) \in \gg$ *denotes that* a *attacks* b*.*

An abstract argumentation framework can be represented as a directed graph, where each vertex is an argument, and a directed edge from a to b if a attacks b. An argumentation framework is acyclic if the corresponding directed graph is acyclic.

For the purposes of this chapter, a semantics maps an argumentation framework to a set of extensions, each extension being a set of arguments (the set of arguments accepted in that extension)[5]. When representing the state of an argument in an extension, we will use the labelling approach (see, for example, [13; 12]) in which the argument is labelled either `in`, `out`, or `undec`. That is, an extension E is defined as a function $\mathcal{L}ab_E : \mathcal{A} \to \{\mathtt{in}, \mathtt{out}, \mathtt{undec}\}$. Then we can define an extension E as $\{a \in \mathcal{A} \mid \mathcal{L}ab_E(a) = \mathtt{in}\}$.

Given an argumentation framework $\mathcal{AF} = (\mathcal{A}, \gg)$, an argument a is said to be *accepted* in an extension E if $\mathcal{L}ab_E(a) = \mathtt{in}$, *rejected* in E if $\mathcal{L}ab_E(a) = \mathtt{out}$, and *undecided* in E if $\mathcal{L}ab_E(a) = \mathtt{undec}$. An extension E is *conflict-free* if no accepted argument is attacked by an accepted argument. An argument a is *defended* by E if every argument that attacks a is attacked by some argument accepted in E. An extension E of $\mathcal{AF}$ is *stable* if it is conflict-free and for every argument $a \in \mathcal{A} \backslash E$

[5] Thus we will not address the gradual and ranking semantics discussed in [15; 1].

there is an argument in E that attacks a. An extension E of $\mathcal{AF}$ is *complete* if it is conflict-free and, $a \in E$ iff a is defended by E.

The set of complete extensions forms a lower semi-lattice under the containment ordering, and many semantics can be defined directly in terms of this semi-lattice. The least complete extension under the containment ordering exists and is called the *grounded* extension. The *preferred* extensions are the maximal complete extensions under the containment ordering. The *semi-stable* extensions are the complete extensions where the set of arguments labelled with `in` or `out` is maximal under the containment ordering. The *ideal* extension is the maximal complete extension contained in all preferred extensions. Similarly, the *eager* extension is the maximal complete extension contained in all semi-stable extensions. These are not necessarily the original definitions of these extensions, but they are equivalent definitions.

We will use $\mathcal{GR}$ to denote the grounded semantics, $\mathcal{ST}$ for the stable semantics, $\mathcal{CO}$ for the complete semantics, $\mathcal{PR}$ for the preferred semantics, $\mathcal{ST}$ for the stable semantics, $\mathcal{SST}$ for the semi-stable semantics, $\mathcal{EA}$ for the eager semantics, and $\mathcal{ID}$ for the ideal semantics.

We say a semantics is *completist* if every argumentation framework is mapped to a set of complete extensions. These semantics will be our main focus. A semantics is *strongly completist* if it is completist and the set of extensions is determined by the semi-lattice structure of the complete extensions. Among the completist semantics are the grounded, preferred, stable, semi-stable, ideal, eager, and complete semantics. All except the stable semantics are strongly completist. Stable extensions are defined by a property of the individual extension, rather than by a structural property within the semi-lattice of complete extensions, and it turns out there is no equivalent structural definition [90]. Stable semantics is also exceptional in that some argumentation frameworks have no stable extensions.

Each semantics implicitly expresses a criterion for what arguments can coherently be accepted together, given an argumentation framework. Each extension in the semantics represents a "reasonable" adjudication, according to that criterion, of the arguments in the argumentation framework.

Our restriction to completist semantics is, then, an implicit requirement that reasonable adjudications are conflict-free, defend all the accepted arguments, and accept all the defended arguments.[6] Each of the semantics, except (obviously) the complete semantics, imposes extra requirements, reflecting different emphases: the grounded semantics is highly sceptical, requiring a minimal set of accepted arguments[7]; the preferred semantics requires maximal sets of accepted arguments;

[6] However, we make this restriction in this chapter only for simplicity, and not on the basis that this implicit requirement is justified.

[7] Or, equivalently, accepting only arguments that are accepted in all complete extensions.

the stable semantics requires that no argument is left undecided; the semi-stable semantics requires minimal sets of undecided arguments; the ideal semantics requires accepting only arguments that are accepted in all preferred extensions, and accepting as many of these as possible; the eager semantics requires accepting only arguments that are accepted in all semi-stable extensions, and accepting as many of these as possible.

The grounded, ideal and eager semantics are *unitary*: they contain exactly one extension. Such semantics limit, somewhat, the range of possible strategic aims of players in strategic argumentation, as we will see later.

Structural properties of an argumentation framework can influence the relationship between various semantics, which can make proving the computational complexity of some problems easier. An argumentation framework is *well-founded* if there is no infinite sequence of arguments $a_1, a_2, \ldots, a_i, a_{i+1}, \ldots$ such that, for each i, a_{i+1} attacks a_i. Such argumentation frameworks have a single complete extension which must be the grounded extension [41], in which every argument is either accepted or rejected. Every completist semantics for such argumentation frameworks consists of this single extension.

An argument framework is *coherent* if every preferred extension is stable. An argument b *indirectly attacks* an argument a if there is a path of odd length from b to a, and *indirectly defends* a if there is a path of even length from b to a. An argument b is *controversial* wrt a if b indirectly attacks a and indirectly defends a. An argument is *controversial* if it is controversial wrt some argument. An argument framework is *uncontroversial* if there is no controversial argument. An argument framework is *limited controversial* if there is no infinite sequence of arguments $a_1, a_2, \ldots, a_i, \ldots$ such that a_{i-1} is controversial wrt a_i. Dung shows that (Theorem 33 of [41]) every limited controversial argument framework is coherent, and every uncontroversial argument framework is also relatively grounded. An argument framework is *relatively grounded* if intersection of all preferred extensions coincides with the grounded extension.

3.3 Structured Argumentation

Argumentation takes place over a language of expressions, most commonly a language of literals. For definiteness, in this chapter we consider propositional literals.

Definition 3.2 (Language)**.** *The* language L *of expressions consists of a set of literals. Given a set* PROP *of propositional atoms, the set of* literals *is* $\text{Lit} = \text{PROP} \cup \{\neg p \mid p \in \text{PROP}\}$. *We denote with* $\sim p$ *the* complementary *of literal* p*; if* p *is a positive literal* q*, then* $\sim p$ *is* $\neg q$*, and if* p *is a negative literal* $\neg q$*, then* $\sim p$ *is* q*.*

Rules are built out of these expressions. Rules have labels to name them, but

these are completely separate from labels used in abstract argumentation.

Definition 3.3 (Rules)**.** *Let* Lab *be a set of rule labels. A* rule r *with* $r \in$ Lab *describes the relation between a set of expressions, called the* antecedent *(or* body *or the* premise*) of* r *and denoted by* $A(r)$ *(which may be empty) and an expression, called the* consequent*, or* head*, of* r *and denoted by* $C(r)$*. Three kind of rules are allowed:* strict rules *of the form* $r : A(r) \to C(r)$*,* defeasible rules *of the form* $r : A(r) \Rightarrow C(r)$*, and* defeaters *of the form* $r : A(r) \rightsquigarrow C(r)$*.*

A strict rule is a rule in the classical sense: whenever the antecedent holds, so does the conclusion. We call a strict rule without antecedent a *fact*, but we often distinguish facts from "true" strict rules that have an antecedent. A defeasible rule is allowed to assert its conclusion unless there is contrary evidence to it. A defeater is a rule that cannot be used to draw any conclusion, but can provide contrary evidence to complementary conclusions. A defeater in this sense [102] can be considered an instance of the general notion of defeater in epistemology: evidence that counts against a belief.

Definition 3.4 (Argumentation Theory)**.** *An* argumentation theory D *is a structure* $(R, >)$*, where* R *is a (finite) set of rules and* $> \subseteq R \times R$ *is a binary relation on* R *called the* superiority relation*.*

The relation $>$ describes the relative strength of rules, that is to say, when a single rule may override the conclusion of another rule, and is required to be irreflexive, asymmetric, and acyclic (i.e., its transitive closure is irreflexive). To simplify discussion, we assume no strict rule is inferior to another rule. We use the following abbreviations on R: the set of strict rules in R is denoted by R_s, the set of strict and defeasible rules in R by R_{sd}, the set of defeasible rules by R_d, the set of defeaters by R_{dft}, and $R[q]$ is the set of rules in R whose head is q.

To demonstrate these definitions, we look at a time-honoured example of defeasible reasoning.

Example 3.5. *Consider an argumentation theory consisting of the following rules*

$$\begin{array}{rrcl}
r_1: & bird(X) & \Rightarrow & fly(X) \\
r_2: & penguin(X) & \Rightarrow & \neg fly(X) \\
r_3: & penguin(X) & \to & bird(X) \\
r_4: & injured(X) & \rightsquigarrow & \neg fly(X) \\
f: & penguin(tweety) & & \\
g: & bird(freddie) & & \\
h: & injured(freddie) & &
\end{array}$$

and a priority relation $r_2 > r_1$.

Here r_1, r_2, r_3, r_4, f *are labels and* r_3 *is (a reference to) a strict rule, while* r_1 *and* r_2 *are defeasible rules,* r_4 *is a defeater, and* f, g, h *are facts. Thus* $R_s = \{r_3, f, g, h\}$ *and* $R_{sd} = R = \{r_1, r_2, r_3\}$ *and* $>$ *consists of the single tuple* (r_2, r_1). *The rules express that birds usually fly* (r_1), *penguins usually don't fly* (r_2), *that all penguins are birds* (r_3), *and that an injured animal may not be able to fly* (r_4). *In addition, we are given the facts that tweety is a penguin, and* $freddie$ *is an injured bird. Finally, the priority of* r_2 *over* r_1 *expresses that when something is both a bird and a penguin (that is, when both rules can fire) it usually cannot fly (that is, only* r_2 *may fire, it overrules* r_1*).*

By combining the rules in a theory, we can build arguments (we adjust the definition in [112] to meet Definition 3.4). In what follows, for a given argument A, Conc returns its conclusion, Sub returns all its sub-arguments, Rules returns all the rules in the argument and, finally, TopRule returns the last inference rule in the argument.

Definition 3.6 (Argument)**.** *Let* $D = (R, >)$ *be an argumentation theory and* $\Rrightarrow \in \{\rightarrow, \Rightarrow, \rightsquigarrow\}$. *An argument* A *constructed from* D *has the form* $A_1, \ldots, A_n \Rrightarrow_r \psi$, *where*

- A_k *is an argument constructed from* D, *for* $1 \leq k \leq n$, *and*
- $r : \mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rrightarrow \psi$ *is a rule in* R.

The set of arguments constructed from D *is the smallest set of arguments satisfying this condition.*

With regard to argument A, *the following holds:*

$$\begin{aligned}
&\mathsf{Conc}(A) = \psi \\
&\mathsf{Sub}(A) = \mathsf{Sub}(A_1) \cup \cdots \cup \mathsf{Sub}(A_n) \cup \{A\} \\
&\mathsf{TopRule}(A) = r : \mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rrightarrow \psi \\
&\mathsf{Rules}(A) = \mathsf{Rules}(A_1) \cup \cdots \cup \mathsf{Rules}(A_n) \cup \{\mathsf{TopRule}(A)\} \\
&(\mathsf{Rules}(A_1) \cup \cdots \cup \mathsf{Rules}(A_n)) \cap R_{dft} = \emptyset
\end{aligned}$$

If $\mathsf{Rules}(A) \subseteq R_s$ *then argument* A *is* strict, *otherwise* A *is* defeasible. *If* $\mathsf{Rules}(A) \cap R_{dft} \neq \emptyset$ *then argument* A *is* non-supportive, *otherwise it is* supportive.

Conflicts between contradictory argument conclusions are resolved on the basis of preferences over arguments using a simple last-link ordering. An argument A is stronger than another argument B (written $A > B$) iff B is defeasible, and either A is strict or $\mathsf{TopRule}(A)$ is stronger than $\mathsf{TopRule}(B)$ ($\mathsf{TopRule}(A) > \mathsf{TopRule}(B)$).

Definition 3.7 (Attacks)**.** *An argument B attacks an argument A iff $\exists A' \in \mathsf{Sub}(A)$ such that $\mathsf{Conc}(B) = \sim\mathsf{Conc}(A')$, and $A' \not> B$.*

We can now define the argumentation framework that is determined by an argumentation theory.

Definition 3.8 (AF determined by an argumentation theory)**.** *Let $D = (R, >)$ be an argumentation theory. The* argumentation framework determined by D *is $(\mathcal{A}, \gg)$, where $\mathcal{A}$ is the set of all arguments constructed from D, and $\gg$ is the attack relation defined above.*

Given this definition of argumentation framework, if D is an argumentation theory, we can abuse notation somewhat and write $\mathcal{GR}(D)$ to denote the grounded extension of the argumentation framework determined by D.

Definition 3.9 (Justified Conclusion)**.** *Given an argumentation theory D, we say a conclusion ψ is* justified by D *under the grounded semantics iff there exists a supportive argument a in $\mathcal{GR}(D)$ such that $\mathsf{Conc}(a) = \psi$.*

The following example illustrates the notions just introduced.

Example 3.10. *Using the rules from Example 3.5, we have arguments:*

$$\begin{array}{lll} A_1: & \rightarrow_f penguin(tweety) & (\textit{strict argument}) \\ A_2: & A_1 \rightarrow_{r_3} bird(tweety) & (\textit{strict argument}) \\ A_3: & A_2 \Rightarrow_{r_1} fly(tweety) & (\textit{defeasible argument}) \\ A_4: & A_1 \Rightarrow_{r_2} \neg fly(tweety) & (\textit{defeasible argument}) \end{array}$$

among others.

If we consider the argument A_3, we have

$$\begin{array}{ll} \mathsf{Conc}(A_3) & = fly(tweety) \\ \mathsf{Sub}(A_3) & = \{A_1, A_2, A_3\} \\ \mathsf{TopRule}(A_3) & = r_1 \\ \mathsf{Rules}(A_3) & = \{f, r_1, r_3\} \end{array}$$

A_4 attacks A_3 because the two arguments have contradictory conclusions and $r_1 \not> r_2$. On the other hand, A_3 does not attack A_4 because $r_2 > r_1$.

In the argumentation framework determined by this theory there is no argument attacking A_4. Hence A_4 appears in the grounded extension. Since A_4 is a supportive argument, its conclusion $\neg fly(tweety)$ is justified under the grounded semantics.

4 Defeasible Logic

Defeasible Logic (DL) [103] is a rule-based sceptical approach to non-monotonic reasoning. It is based on a logic programming-like language and is a simple, efficient but flexible formalism capable of dealing with many intuitions of non-monotonic reasoning in a natural and meaningful way [4].

Defeasible rule languages like defeasible logic have been shown to be useful in representing legal documents and reasoning [113; 9; 118; 68; 66; 74; 72]. There are a variety of defeasible logics, which have been argued to represent the different proof standards that apply in legal systems [62; 64].

Defeasible logics have much in common with argumentation, but there is only little work substantiating the relationship. [65] characterizes inference in two defeasible logics in terms of argumentation. [62] maps proof in Carneades [59] at a given proof standard into proof in a defeasible logic. [79] showed how to map one instance of $ASPIC^+$ into a defeasible logic. [93] gave two embeddings of abstract argumentation frameworks $\mathcal{AF}$ into a small subset of defeasible rule languages, implying, in particular, that acceptance in the grounded extension of $\mathcal{AF}$ can be implemented in a wide variety of defeasible logics and other concrete defeasible reasoning formalisms.

In this section we define two defeasible logics, but first we introduce defeasible logic in general.

4.1 Defeasible logic

The language of DL consists of literals and rules. To avoid notational redundancies, we use the same definitions of PROP, Lit, complementary literal, and the same rule types, structure and notation as already introduced in Definition 3.2.

A defeasible theory D is a triple $(F, R, >)$, where $F \subseteq \text{Lit}$ is a set of indisputable statements called *facts*, R is a (finite) set of rules, and $> \subseteq R \times R$ is a superiority relation on R as introduced in Definition 3.4.

A *derivation* (or *proof*) is a finite sequence $P = P(1), \ldots, P(n)$ of *tagged literals* of the type $+\Delta q$ (q is definitely provable), $-\Delta q$ (q is definitely refuted), $+d\,q$ (q is defeasibly provable) and $-d\,q$ (q is defeasibly refuted). The proof conditions below define the logical meaning of such tagged literals. Given a proof P, $P(n)$ denotes the n-th element of the sequence, and $P(1..n)$ denotes the first n elements of P. $\pm\Delta$ and $\pm df$ are called *proof tags*. Given $\#$ a proof tag, the notation $D \vdash \pm\#q$ means that there is a proof P in D such that $P(n) = \pm\#q$ for an index n.

In the remainder, we only present the proof conditions for the positive tags: 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 [5].

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

$+\Delta$: If $P(n+1) = +\Delta q$ then either
(1) $q \in F$ or
(2) $\exists r \in R_s[q]$ s.t. $\forall a \in A(r).\ +\Delta a \in P(1..n)$.

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. Literal q is definitely refuted if (1) is not a fact and (2) every strict rule for q has at least one definitely refuted antecedent. Conceptually, strict derivations are much stronger than defeasible ones: the superiority relation plays no part in them. If we have two strict rules for opposite conclusions whose antecedents are all proven, then the logic will derive both conclusions, which signals an inconsistency within the theory itself.

The conditions to establish a defeasible proof $+d$ have a structure similar to arguments, and are formalised by the following schema.

$+d$: If $P(n+1) = +d\,q$ then either
(1) $+\Delta q \in P(1..n)$ or
(2) (2.1) $-\Delta{\sim}q \in P(1..n)$ and
(2.2) $\exists r \in R_{sd}[q]$ s.t. r is applicable, and
(2.3) $\forall s \in R[{\sim}q]$. either
(2.3.1) s is unsupported, or
(2.3.2) s is defeated.

Intuitively, a rule is applicable if all the literals in the antecedent have previously been proven. Clause (2.3) considers the possible counter-arguments. To derive q, each such counter-argument must be either unsupported, or defeated. A rule is unsupported if it is not possible to give a (valid) justification for at least one of the premises of the rule. The degree of provability of the conclusion we want to obtain determines the meaning of valid justification for a premise. This could vary from a derivation for the premise to a simple chain of rules leading to it. Finally, a rule is defeated if there is an applicable rule stronger than it.

By instantiating the abstract definitions of applicable, supported and defeated, the above structure defines several variants of DL. In particular, we address the distinction between *ambiguity blocking* and *ambiguity propagation*. A literal q is *ambiguous* if (i) there is a chain of reasoning that supports a conclusion q, (ii) one (chain) supporting the complementary conclusion ${\sim}q$, and (iii) the superiority relation does not resolve this conflict.

Example 4.1. *Consider the defeasible theory* $D = (\emptyset, R, \emptyset)$*, such that*

$$R = \{r_1 : \Rightarrow a, \quad r_2 : \Rightarrow b, \quad r_3 : \Rightarrow \neg a, \quad r_4 : a \Rightarrow \neg b\}.$$

Here a *is ambiguous since both* r_1 *and* r_3 *are applicable, and there is no superiority between them.*

In what follows we shall introduce two variants of DL, the first one supporting ambiguity blocking, and the second one supporting ambiguity propagation. We explain the intuitions behind the two variants by referring to Example 4.1, where a is ambiguous. In a setting where ambiguity is blocked, b is not ambiguous because rule r_2 for b is applicable, whilst r_4 for $\neg b$ is not, since we cannot prove a. On the other hand, in an ambiguity propagating setting, b is ambiguous because a is not disproved, and so the applicability of r_4 is not denied. In this way, the ambiguity is propagated to b.

The ambiguity blocking and ambiguity propagation is a clash in intuitions in non-monotonic reasoning [130]. However, [62] argues that the distinction can be used to characterise different proof standards, where ambiguity blocking corresponds to the proof standard of *preponderance of evidence* while ambiguity propagation captures the *beyond reasonable doubt* proof standard. Furthermore, there are scenarios where both intuitions are needed (for different conclusions), and the reasoning for conclusions requiring one of the two proof standard depends on conclusions obtained using the other proof standard. See [64] for the details and how to combine the two intuitions.

In the remainder, we shall use ∂ for the proof tag to indicate that a conclusion is defeasibly provable (refutable) under ambiguity blocking, and δ for the corresponding notions under ambiguity propagation.

4.2 Ambiguity Blocking Defeasible Logic

The ambiguity blocking variant of DL was introduced in [7] and is captured by the following instantiation of $+d$:

$+\partial$: If $P(n+1) = +\partial q$ then either
(1) $+\Delta q \in P(1..n)$ or
(2) (2.1) $-\Delta{\sim}q \in P(1..n)$ and
(2.2) $\exists r \in R_{sd}[q]$ s.t. $\forall a \in A(r) + \partial a \in P(1..n)$ and
(2.3) $\forall s \in R[{\sim}q]$ either
(2.3.1) $\exists a \in A(s)$ s.t. $-\partial a \in P(1..n)$ or
(2.3.2) $\exists t \in R_{sd}[q]$ s.t.
$\forall a \in A(t) + \partial a \in P(1..n)$ and $t > s$.

To prove $+\partial q$, we have to show that either (1) q is already definitely provable, or (2.2) there is an applicable rule for q and (2.3) for very rule attacking q either (2.3.1) at least one antecedent has been defeasibly refuted, or (2.3.2) the rule is defeated by a (stronger) rule for q.

In other terms, a rule is applicable if all the elements of the body are defeasibly provable. A rule is unsupported if there is an element of the body that is defeasibly refuted. A rule is defeated if it is weaker than an applicable rule. We use DL(∂) to denote the ambiguity blocking defeasible logic variant.

4.3 Ambiguity Propagating Defeasible Logic

Ambiguity propagation describes a behaviour where ambiguity of a literal is propagated to dependent literals. This is achieved in DL by separating the invalidation of a counterargument from the derivation of tagged literals. To do so, another kind of conclusion, called *support* and denoted by Σ, is introduced [8].

$+\Sigma$: If $P(n+1) = +\Sigma q$ then either
(1) $+\Delta q \in P(1..n)$ or
(2) (2.1) $-\Delta{\sim}q \in P(1..n)$ and
(2.2) $\exists r \in R_{sd}[q]$ s.t.
(2.2.1) $\forall a \in A(r) + \Sigma a \in P(1..n)$ and
(2.2.2) $\forall s \in R[{\sim}q]$ either
$\exists a \in A(s)$ s.t. $-\delta a \in P(1..n)$, or $s \not> r$.

The condition for $+d$ is thus instantiated as follows:

$+\delta$: If $P(n+1) = +\delta q$ then either
(1) $+\Delta q \in P(1..n)$ or
(2) (2.1) $-\Delta{\sim}q \in P(1..n)$ and
(2.2) $\exists r \in R_{sd}[q]$ s.t. $\forall a \in A(r) + \delta a \in P(1..n)$ and
(2.3) $\forall s \in R[{\sim}q]$ either
(2.3.1) $\exists a \in A(s)$ s.t. $-\Sigma a \in P(1..n)$ or
(2.3.2) $\exists t \in R_{sd}[q]$ s.t.
$\forall a \in A(t) + \delta a \in P(1..n)$ and $t > s$.

The idea is that a conclusion q is supported if (2.1) there is a rule for q such that (2.2.1) all the elements in the antecedent are (at least) supported, and that (2.2.2) all rules for the opposite conclusion have (at least) one premise that has been refuted, or such a rule is not stronger than the rule for q. This means that there is an undefeated argument supporting the conclusion. Then to affirm that a conclusion is provable,

we have to provide an argument/rule where all the antecedents are provable, and there is no argument/rule for the opposite that is at least supported. We refer to the ambiguity propagating variant by using DL(δ).

Example 4.1 (Continued)**.** *Consider, again, the theory* $D = (\emptyset, R, \emptyset)$*, where*

$$R = \{r_1 : \Rightarrow a, \quad r_2 : \Rightarrow b, \quad r_3 : \Rightarrow \neg a, \quad r_4 : a \Rightarrow \neg b\}.$$

By definition of $+\partial$*, we obtain the following conclusions from* D*:* $-\partial a$*,* $-\partial\neg a$*,* $+\partial b$*,* $-\partial\neg b$*, capturing the ambiguity blocking behaviour of* DL(∂)*. On the other hand, if we compute the consequences of* D *by using the proof conditions for* Σ *and* δ*, we obtain* $+\Sigma a$*,* $+\Sigma\neg a$*,* $+\Sigma b$*,* $+\Sigma\neg b$ *and thus also* $-\delta a$*,* $-\delta\neg a$*,* $-\delta b$ *and* $-\delta\neg b$*. In this way, we capture the ambiguity propagation feature of* DL(δ)*.*

4.4 Team or Individual Defeat?

The defeasible logics defined above have the property of *team defeat*: the rules for a literal q are compared with the rules for $\sim q$. If each applicable rule for $\sim q$ is inferior to some applicable rule for q, then the rules for q, as a team, overcome the rules for $\sim q$. Thus, q is inferred. In comparison, under *individual defeat* there must be an applicable rule for q that is superior to *all* applicable rules for $\sim q$ in order to overcome the rules for $\sim q$ and infer q. Clearly, any time individual defeat overcomes the rules for $\sim q$, so does team defeat.

To get some intuition about these two forms of defeat we use a variation of an example from [7].

Example 4.2. *Consider some rules of thumb about animals and, particularly, mammals. An egg-laying animal is generally not a mammal. Similarly, an animal with webbed feet is generally not a mammal. On the other hand, an animal with fur is generally a mammal. Finally, the monotremes are a subclass of mammal. These rules are represented as defeasible rules below.*

Furthermore, animals with fur and webbed feet are generally mammals, so r_2 *should overrule* r_4*. And monotremes are a class of egg-laying mammals, so* r_1 *should overrule* r_3*.*

Finally, it happens that a platypus is a furry, egg-laying, web-footed monotreme. Is it a mammal? (That is, is $mammal(platypus)$ *a consequence of the defeasible theory below?)*

$r_1 : \ monotreme(X) \Rightarrow mammal(X)$ $\quad$ $r_3 : \ laysEggs(X) \ \Rightarrow \neg mammal(X)$
$r_2 : \ hasFur(X) \ \Rightarrow mammal(X)$ $\quad$ $r_4 : \ webFooted(X) \Rightarrow \neg mammal(X)$
$r_1 > r_3$ $\quad$ $r_2 > r_4$

$monotreme(platypus)$ $\quad$ $laysEggs(platypus)$
$hasFur(platypus)$ $\quad$ $webFooted(platypus)$

It is obvious that all four rules are applicable to the question of $mammal(platypus)$. *Under team defeat, each rule for* $\neg mammal(platypus)$ *is overcome by some rule for* $mammal(platypus)$, *so* $mammal(platypus)$ *is inferred. However, there is no single rule for* $mammal(platypus)$ *that overcomes all rules for* $mammal(platypus)$, *so under individual defeat we cannot infer* $mammal(platypus)$ *(nor* $\neg mammal(platypus)$*).*

Thus, we see that team defeat can be useful in making a justified inference that otherwise would not be made. On the other hand, most expressions of structured argumentation employ individual defeat.

Fortunately, it is easy to adjust the inference conditions for the two logics defined above to obtain individual defeat: we simply replace the sub-conditions (2.3.2) by $r > s$. We denote the individual defeat logics by DL(∂^*) and DL(δ^*). For more discussion of the four variants of defeasible logic discussed here, see [23].

Finally, we consider the relationship between these logics. A series of papers [84; 85; 86; 87] investigates the relative expressiveness of variants of Defeasible Logic. In brief, two (defeasible) logics L_1 and L_2 have the same expressiveness iff the two logics simulate each other (where a defeasible logic L_2 simulates a defeasible logic L_1 if there is a polynomial time transformation T that transforms a theory D_1 of L_1 in a theory $D_2 = T(D_1)$ of L_2 such that, for any addition of facts A, all strict and defeasible conclusions of $D_1 \cup A$ are the same as those of $D_2 \cup A$ in L_1). [84; 85] provide polynomial time transformations between each of the four logics defined above.

Theorem 4.3. *[85] Each of* DL(∂), DL(δ), DL(∂^*), *and* DL(δ^*) *simulates the others.*

5 Strategic Argumentation for Defeasible Logic and Structured Argumentation

We now propose a Defeasible Logic instantiation of the games introduced in Section 2. We shall hence specialise Definitions 2.6 and 2.7 for the instance at hand, and then proceed with the formulation of two problems.

Given a defeasible theory $D = (F, R, >)$, we define the corresponding *split defeasible theory* as $SD = (F_{\mathsf{Com}}, F_{\mathsf{Pr}}, F_{\mathsf{Op}}, R_{\mathsf{Com}}, R_{\mathsf{Pr}}, R_{\mathsf{Op}}, >)$ with $F = F_{\mathsf{Com}} \cup F_{\mathsf{Pr}} \cup F_{\mathsf{Op}}$ and $R = R_{\mathsf{Com}} \cup R_{\mathsf{Pr}} \cup R_{\mathsf{Op}}$. We call the content of dispute discussed by the players the *critical literal*, and note that the arguments brought about by the players

will be in the form of defeasible derivations. We assume that each player is informed about the restriction of $>$ to their private rules,

We will have three instances of the definitions of Section 2, owing to the extra expressivity of defeasible logic. Defeasible logic offers the following three ways to express a contrary to $D \vdash +d\,q$: the negation of q can be proved ($D \vdash +d\,{\sim}q$); within the logic we can prove that that $+d\,q$ cannot be proved ($D \vdash -d\,q$); and, we cannot prove $+d\,q$ ($D \nvdash +d\,q$). Thus, if Pr wants to prove q, Op has three possible levels of opposition. The first will lead to a symmetric game, and the third to an asymmetric game. The second falls somewhere in between, and we will call it a *semi-symmetric* game. In the semi-symmetric game Op shoulders a burden of proof, but only to prove that Pr's aim cannot be proved, not to prove the negation of q.

If we consider the asymmetric case corresponds to the Scottish verdict of not proven[8] and the symmetric case corresponds to not guilty, then what is the semi-symmetric case? Technically, in defeasible logic, the distinction between semi-symmetric and asymmetric opposition is caused by a circularity or infinite regress in an argument. Abstractly, it might represent unknowability, or an incapacity of the proceedings/inference rules – inability to decide that l is not provable, even though l, in fact, is not provable (a little bit like Gödel's incompleteness theorem).

The game rules discussed in Section 2 are instantiated as follows. The parties start the game by choosing the critical literal l. Pr has the burden to prove $+d\,l$ by using the remainder of its private rules along with those that currently have been played; Op's final aim is to prove $+d\,{\sim}l$ in the symmetric version of the game, to prove $-d\,l$ in the semi-symmetric game, and simply to prevent the proof of $+d\,l$ in the asymmetric game.

Note that, when putting forward an argument, the players: (1) may propose, along with a subset of their private rules, a subset of their private facts to support such rules (see Example 5.2 at the end of this section), and (2) may play an argument whose terminal literal differs from l or ${\sim}l$ (with the aim to attack/disprove one of the premises of a rule in the proof proving $l/{\sim}l$).

As the semi-symmetric and asymmetric games differ from the symmetric one only in Op's final aim, to avoid pedantic redundancies we shall provide a single definition for the three games.

Definition 5.1 (SSA (SSSA, AsSA) Game for Defeasible Logic)**.** *Consider two players, a proponent* Pr *and an opponent* Op*, a split defeasible theory* $SD =$

[8] Roughly, under this verdict the jury considers the prosecution has not made the case for "guilty", beyond a reasonable doubt, but the defence has not made the case for "innocent". A verdict of *guilty* is given when the jury considers the prosecution has made its case, and *not guilty* when the defence has made its case. See [11] or the Wikipedia entry for *Not proven*.

$(F_{\mathsf{Com}}, F_{\mathsf{Pr}}, F_{\mathsf{Op}}, R_{\mathsf{Com}}, R_{\mathsf{Pr}}, R_{\mathsf{Op}}, >)$, *and a critical literal* $l \in L$.

Let F^k_{Com}, R^k_{Com}, F^k_{Pr}, R^k_{Pr}, F^k_{Op}, *and* R^k_{Op} *denote, respectively, the common (knowledge) facts and rules,* Pr*'s private facts and rules, and* Op*'s private facts and rules, after turn* k. *(In particular,* $F^0_{\mathsf{Com}} = F_{\mathsf{Com}}$, $R^0_{\mathsf{Com}} = R_{\mathsf{Com}}$, $F^0_{\mathsf{Pr}} = F_{\mathsf{Pr}}$ $R^0_{\mathsf{Pr}} = R_{\mathsf{Pr}}$, $F^0_{\mathsf{Op}} = F_{\mathsf{Op}}$, *and* $R^0_{\mathsf{Op}} = R_{\mathsf{Op}}$.*) The* common defeasible theory *at that point is* $D^k = (F^k_{\mathsf{Com}}, R^k_{\mathsf{Com}}, >)$.

We define a symmetric (resp. semi-symmetric, asymmetric) strategic argumentation game for Defeasible Logic *as a dialogue game where:*

1. *The players take turns. If* $D^0 \vdash +d\, l$ *then* Op *begins; otherwise* Pr *does so.*
2. *At turn* k, *if* $D^{k-1} \vdash +d\, \neg l$ *(resp.* $D^{k-1} \vdash -d\, l$ *for the semi-symmetric version,* $D^k \not\vdash +d\, l$ *for the asymmetric version), then it is* Pr*'s turn to play, as follows*
   - Pr *advances a subset of its private facts* $\Phi \subseteq F^{k-1}_{\mathsf{Pr}}$ *and rules* $\rho \subseteq R^{k-1}_{\mathsf{Pr}}$ *so that* $D^k \vdash +d\, l$. *As a result*
     - $F^k_{\mathsf{Com}} = F^{k-1}_{\mathsf{Com}} \cup \Phi$ *and* $R^k_{\mathsf{Com}} = R^{k-1}_{\mathsf{Com}} \cup \rho$;
     - $F^k_{\mathsf{Pr}} = F^{k-1}_{\mathsf{Pr}} \setminus \Phi$ *and* $R^k_{\mathsf{Pr}} = R^{k-1}_{\mathsf{Pr}} \setminus \rho$;
     - $R^k_{\mathsf{Op}} = R^{k-1}_{\mathsf{Op}}$.
3. *At turn* k, *if* $D^{k-1} \vdash +d\, l$, *then it is* Op*'s turn to play, as follows*
   - Op *advances a subset of its private* $\Phi \subseteq F^{k-1}_{\mathsf{Op}}$ *and rules* $\rho \subseteq R^{k-1}_{\mathsf{Op}}$ *so that* $D^k \vdash +d\, \neg l$ *(resp.* $D^k \vdash -d\, l$ *for the semi-symmetric version,* $D^k \not\vdash +d\, l$ *for the asymmetric version). As a result*
     - $F^k_{\mathsf{Com}} = F^{k-1}_{\mathsf{Com}} \cup \Phi$ *and* $R^k_{\mathsf{Com}} = R^{k-1}_{\mathsf{Com}} \cup \rho$;
     - $R^k_{\mathsf{Pr}} = R^{k-1}_{\mathsf{Pr}}$;
     - $F^k_{\mathsf{Op}} = F^{k-1}_{\mathsf{Op}} \setminus \Phi$ *and* $R^k_{\mathsf{Op}} = R^{k-1}_{\mathsf{Op}} \setminus \rho$.
4. *The game ends at turn* $k+1$, *when either (i) it is* Pr*'s turn and there is no move for* Pr *such that the common defeasible theory* $D^{k+1} \vdash +d\, l$, *in which case* Op *wins, or (ii) it is* Op*'s turn and there is no move for* Op *such that the common defeasible theory* $D^{k+1} \vdash +d\, \neg l$ *(resp.* $D^{k+1} \vdash -d\, l$ *for the semi-symmetric version,* $D^k \not\vdash +d\, l$ *for the asymmetric version), in which case* Pr *wins.*

The corresponding decision problems are as follows.

SSA (SSSA, AsSA) Problem for Defeasible Logic

Let SD^k be a split defeasible theory as in Definition 5.1 after turn k, D^{k+1} be the corresponding common defeasible theory after turn $k+1$, and $l \in L$ be the critical literal.

Pr's instance for turn $k+1$: Let F^k_{Pr} and R^k_{Pr} be, respectively, the set of Pr's private facts and rules after turn k, and that the common defeasible theory assume $D^k \vdash +d\,\neg l$ (resp. $D^k \vdash -d\,l$ and $D^k \nvdash +d\,l$ for the semi-symmetric and asymmetric problems).

Question: Do there exist Φ subset of F^k_{Pr} and ρ subset of R^k_{Pr} such that the common defeasible theory $D^{k+1} \vdash +d\,l$?

Op's instance for turn $k+1$: Let F^k_{Op} and R^k_{Op} be, respectively, the set of Op's private facts and rules after turn k, and assume that the common defeasible theory $D^k \vdash +d\,l$.

Question: Do there exist Φ subset of F^k_{Op} and ρ subset of R^k_{Op} such that the common defeasible theory $D^{k+1} \vdash +d\,\neg l$ (resp. $D^{k+1} \vdash -d\,l$ and $D^{k+1} \nvdash +d\,l$, for the semi-symmetric and asymmetric problems)?

We explore how these games are played through an example theory that shows how different moves by the players may lead to different result of the game in the symmetric and semi-symmetric/asymmetric variants.

Example 5.2. *Consider* $SD = (F_{\mathsf{Com}}, F_{\mathsf{Pr}}, F_{\mathsf{Op}}, R_{\mathsf{Com}}, R_{\mathsf{Pr}}, R_{\mathsf{Op}}, >)$ *such that*

- $F_{\mathsf{Com}} = \{a\}$ *and* $R_{\mathsf{Com}} = \emptyset$*;*
- $F_{\mathsf{Pr}} = \{d\}$ *and* $R_{\mathsf{Pr}} = \{r_1 : a \Rightarrow p,\ r_2 : b, d \Rightarrow p\}$*;*
- $F_{\mathsf{Op}} = \{b, c\}$ *and* $R_{\mathsf{Op}} = \{r_3 : c \Rightarrow \neg p,\ r_4 : b \Rightarrow \neg p\}$*; and*
- $> \{(r_4, r_1), (r_2, r_4)\}$.

The critical literal is p. Pr *starts the game and can only advance* r_1*; the fact that* b *is not proven makes* r_2 *unsupported. Consequently, for both variants,* $SD^1 \vdash +d\,p$*. We now detail the different scenarios for* Op *wrt the symmetric, semi-symmetric, and asymmetric games.*

Symmetric variant. Op *considers playing* r_3 *but realises that is not a legal move. In fact, as* r_3 *is neither stronger than* r_1 *nor* r_2*, by playing it* Op *would not prove* $+d\,\neg p$*. By playing* r_4*,* Op *must also advance* r_4*'s only premise,* b *(*$SD^2 \vdash +d\,\neg p$ *and* $SD^2 \vdash +d\,b$*). This makes* r_2 *applicable and allows* Pr *to play it and win the game.*

Semi-symmetric variant. *For this variant of the game,* Op *has the burden to prove* $-d\,p$ *and plays, again,* r_4 *(*$SD^2 \vdash +d\,\neg p$ *and* $+d\,\neg p$ *implies* $-d\,p$*).* Pr *can again play* r_2 *leading to* $SD^3 \vdash +d\,p$*, but now if* Op *plays* r_3 *(along with* c*), then* $SD^4 \vdash -d\,p$*.* Pr *has no more rules to play and this time* Op *wins.*

Asymmetric variant. *This variant of the game unfolds in the same way as the semi-symmetric variant because, for every k, $SD^k \vdash -d\,p$ implies $SD^k \nvdash +d\,p$.*

We can modify the above example to demonstrate the distinction between the semi-symmetric and asymmetric games.

Example 5.3. *Consider the modification of Example 5.2 where r_3 in R_{Op} is replaced by*

$$r_3 : c, \neg p \Rightarrow \neg p$$

Symmetric variant. *This variant unfolds in exactly the same way as Example 5.2.* Op *does not play r_3.*

Semi-symmetric variant. *For this variant of the game,* Pr *plays r_1,* Op *plays r_4, and* Pr *plays r_2, just as in the symmetric variant. At this stage* Op *would like to play r_3 but, again, this is not a legal move: playing it would not achieve $SD^4 \vdash -d\,p$. Thus* Pr *wins.*

Asymmetric variant. *Again,* Pr *plays r_1,* Op *plays r_4, and* Pr *plays r_2. However, in this variant* Op *can play r_3, because then $SD^3 \nvdash +d\,p$.* Pr *has no more moves, so* Op *wins. Alternatively,* Op *could simply play r_3 on her first move, to which* Pr *has no response. Thus* Op *wins without exposing r_4 and b (and without inducing* Pr *to expose r_2 and d).*

We end this subsection with a brief discussion of fact-based strategic argumentation [88], a refinement of the strategic argument games where players can only play facts. That is, strategic argument games where $R_{\mathsf{Pr}} = \emptyset$ and $R_{\mathsf{Op}} = \emptyset$. While general strategic argumentation can be a model for legal argumentation in general, this refinement reflects argument about whether regulations have been adhered to. The players are the party subject to the regulations, and the enforcement body for the regulations. R_{Com} represents the regulations, which are fixed. The players can only generate arguments by marshalling facts that support the applicability of clauses in the regulations (i.e. rules) that, in turn, support the player's contentions. This refinement could also be considered a crude partial model for pleadings in civil law (in that it elicits claimed facts from parties), although different in many ways from Gordon's Pleadings Game [58].

Although this refinement appears to simplify the reasoning required to play the game, in one sense it is no simpler [88]. Any general strategic argumentation game $SD = (F_{\mathsf{Com}}, F_{\mathsf{Pr}}, F_{\mathsf{Op}}, R_{\mathsf{Com}}, R_{\mathsf{Pr}}, R_{\mathsf{Op}}, >)$ can be reduced to the "simpler" game as follows: for each rule $r_i : \beta \Rrightarrow \varphi$ in R_{Pr} we add the rule $r_i : \beta, \alpha(r_i) \Rrightarrow \varphi$ to R_{Com} and add the fact $\alpha(r_i)$ to F_{Pr}, where $\alpha(r_i)$ is a new proposition. And similarly for Op. Every move in the resulting game $SD' = (F_{\mathsf{Com}}, F'_{\mathsf{Pr}}, F'_{\mathsf{Op}}, R'_{\mathsf{Com}}, \emptyset, \emptyset, >)$ corresponds to a move of SD, and vice versa.

5.1 Computational Results

We are now ready to show that deciding what arguments to play at a given turn of a dialogue game under Dung's grounded semantics is an NP-complete problem even when the problem of deciding whether a conclusion follows from an argument is computable in polynomial time.

[67] proved that this problem is NP-complete for DL with ambiguity blocking, i.e., DL(∂). We present here an outline of the proof in [88]. Theorem 5.4 is provided from the viewpoint of Pr. The same result holds for Op.

Theorem 5.4. *The SSA Problem under* DL(∂) *is* NP*-complete.*

Proof. First, the SSA Problem is polynomially solvable on non-deterministic machines. Consider a dialogue game with sets R^0_{Com}, R^0_{Pr}, R^0_{Op} and the defeasible theory $D^{i-1} = (\emptyset, R^{i-1}_{\mathsf{Com}}, >)$, the theory at turn $i-1$ of a dialogue game. An oracle guesses a set of rules $R^i \subseteq R^{i-1}_{\mathsf{Pr}}$, we compute the consequences of the argumentation theory $D^i = (\emptyset, R^{i-1}_{\mathsf{Com}} \cup R^i, >)$, and we check whether the critical literal is a positive or negative consequence. The computation of consequences can be done in polynomial time [83; 23].

Second, we reduce 3SAT to the SSA Problem, proving therefore that the problem is NP-hard. Consider a 3SAT formula $\varphi = \bigwedge_{j=1}^{n} C_j$ such that $C_j = \bigvee_{k=1}^{3} x_j^k$. R^i is defined as follows:

1. For each proposition x occurring in φ, R^{i-1}_{Pr} and R^{i-1}_{Op} both contain

$$t_x : \quad \Rightarrow x$$

$$t_{\neg x} : \quad \Rightarrow \neg x.$$

2. For each clause C_j, R^{i-1}_{Com} contains

$$r_j^k : x_j^k \Rightarrow c_j$$

where x_j^k is either a positive literal (x), or a negative literal ($\neg x$).

3. R^{i-1}_{Com} also contains

$$r_{sat} : c_1, \ldots, c_n \Rightarrow sat.$$

For any assignment θ of values to the Boolean variables in φ, let S_θ be the set of x literals that evaluate to true under θ. And for any consistent subset S of x literals, let θ_S be an assignment that evaluates all elements of S to true. We leave it for the reader to verify that if θ satisfies φ then choosing the move S_θ wins for Pr, and if S is a winning move for Pr then S is consistent and θ_S satisfies φ. □

The same result holds for the semi-symmetric and asymmetric games.

Theorem 5.5. *The SSSA and AsSA problems under* DL(∂) *is* NP*-complete.*

Proof. The proof is essentially the same as that of Theorem 5.4 except for the case when, at turn i, Op must play. In that case, the reduction is identical to the one proposed above, with the only difference that Point 3. now also adds to R^i_{Com} the following rule

$$r_{nsat}: \quad \Rightarrow \neg sat$$

It is trivial to prove that an interpretation satisfies φ iff r_{sat} is applicable iff sat and $\neg sat$ are ambiguous. Thus φ is satisfied iff $-\partial\, sat$ is proved iff $\neg sat$ is not proved. □

While it is possible to define DL(∂) in terms of an argumentation semantics, the logic corresponding to Dung's grounded semantics is ambiguity-propagating [65; 79].

The next step is to determine the computational complexity of the problem at hand for the ambiguity propagating variant of DL. The NP-completeness of the strategic argumentation problem under DL(δ^*) follows immediately from Theorems 4.3, 5.4, and 5.5.

Theorem 5.6. *The SSA, SSSA, and AsSA problems under* DL(δ^*) *are* NP*-complete.*

We have the same results for DL(∂^*) and DL(δ).

In [79], it is shown that the conclusions of an $ASPIC^+$ argumentation theory under grounded semantics are the same as those in DL(δ^*) (after minor changes to the superiority relation).

Theorem 5.7. *[79] Given an $ASPIC^+$ argumentation theory AT, there is a defeasible theory $T(AT)$ such that p is derived under the grounded semantics from AT iff $+\delta^* p$ can be derived from $T(AT)$. Furthermore, all consequences of AT can be computed in time polynomial in the size of AT.*

Thus we can use implementations of DL(δ^*) to implement $ASPIC^+$ argumentation theories that employ the last-link ordering of arguments and the grounded semantics.

We can solve the strategic argumentation problem by non-deterministically choosing a set R^i of rules and then verifying whether the critical literal p is justified in the argumentation framework determined by D^i, or not. Further, the literals justified by the grounded semantics are computable in polynomial time, as shown above. The strategic argumentation problem is thus in NP.

Now, from Theorems 5.6 and 5.7, we obtain the following result.

Theorem 5.8. *The strategic argumentation problems under the grounded semantics are* NP*-complete.*

6 Strategic Abstract Argumentation

In this section we look beyond the grounded semantics to a wide range of other semantics for abstract argumentation frameworks. After exploring the range of dialogue games that can be played in the context of abstract argumentation, we investigate the possibilities for player aims, and identify the complexity of two computational problems related to playing strategic abstract argumentation games, for selected aims and semantics.

6.1 Strategic Argumentation in the Abstract

We formulate a split argumentation framework in this abstract sense as a tuple $(\mathcal{A}, \mathcal{A}_{Com}, \mathcal{A}_{\mathsf{Pr}}, \mathcal{A}_{\mathsf{Op}}, \gg)$ where $\mathcal{A}_{Com}$ is a set of abstract arguments that are common knowledge to the players, $\mathcal{A}_{\mathsf{Pr}}$ ($\mathcal{A}_{\mathsf{Op}}$) is the set of arguments known to Pr (Op), and $\gg$ is the attack relation over all arguments. Each player knows $\gg$ restricted to the set of arguments the player knows. For example, Pr knows $\gg$ restricted to $(\mathcal{A}_{Com} \cup \mathcal{A}_{\mathsf{Pr}}) \times (\mathcal{A}_{Com} \cup \mathcal{A}_{\mathsf{Pr}})$. Each player has a *strategic aim* or *desired outcome* (the two terms will be treated as equivalent) that expresses their desired property of the state of the argument framework at the end of the strategic argumentation game.

A strategic abstract argumentation game consists of alternating moves by Pr and Op until one player cannot make a move. In that case the other player wins. Pr starts the game by playing a set of arguments, including a mutually agreed *critical argument* which is the subject of the two players' strategic aims[9]. By "playing a set of arguments" we refer to the transfer of a set of arguments from the player's set of arguments to $\mathcal{A}_{Com}$ such that the revised common argumentation

[9] Aims will be discussed in the next subsection.

framework $(\mathcal{A}_{Com}, \gg)$ satisfies the player's strategic aim. Thus a move by Pr replaces a split argument framework $(\mathcal{A}_{Com}, \mathcal{A}_{\mathsf{Pr}}, \mathcal{A}_{\mathsf{Op}}, \gg)$ by a new framework $(\mathcal{A}_{Com} \cup X, \mathcal{A}_{\mathsf{Pr}} \backslash X, \mathcal{A}_{\mathsf{Op}}, \gg)$, where $X \subseteq \mathcal{A}_{\mathsf{Pr}}$ is the set of arguments played by Pr in that move, and the new framework achieves Pr's strategic aim. Similarly, a move by Op transfers arguments from $\mathcal{A}_{\mathsf{Op}}$ to $\mathcal{A}_{Com}$. Clearly, if $\mathcal{A}_{\mathsf{Pr}}$ or $\mathcal{A}_{\mathsf{Op}}$ is finite then the game terminates. We will only consider games where $\mathcal{A}_{Com}$, $\mathcal{A}_{\mathsf{Pr}}$ and $\mathcal{A}_{\mathsf{Op}}$ are finite.

Thus, a strategic abstract argumentation game is a dialogue game played by two players (Pr and Op). Let *conc* map arguments to distinct propositions, and let φ be the conclusion of the critical argument. Then the game is an asymmetric strategic argumentation game, as defined in Definition 2.7, where "φ is accepted" is defined as: Pr's aim wrt the critical argument is satisfied.

We assume that the players agree on what is an argument, and whether one argument attacks another. This is implicit in the formulation as a split argumentation framework. But, in theory, there is no reason why the two players should employ the same semantics when they play a strategic argumentation game. For example, Pr might formulate her aim in terms of the preferred semantics, while Op's aim is expressed in terms of the eager semantics. Indeed, it is quite reasonable that different players might perceive the world differently. This is no impediment to the players playing a strategic argumentation game, since the definition of the game only describes moves a player may make, and not the interpretation she puts on the game.

However, there has not been any work on such situations. This is not so surprising when we consider that strategic argumentation is primarily treated as an adversarial game. Real world situations that are modelled by strategic argumentation may need the presence of an adjudicator to enforce any conclusions that result from the game. Such an adjudicator might bring their own perceptions and semantics to the game. Thus, playing in a common semantics could be considered as both players adopting the adjudicator's view of the world.

Similarly, there is no *prima facie* reason why the two players should focus on a single critical argument, rather than have individual, separate foci. The literature has rarely addressed this possibility ([71] is an exception). However, once we assume that the players agree on a focus, the use of a single critical argument for each player implies no loss of generality. Straightforward constructions can map a disjunction or conjunctions of arguments to a single argument in most semantics[10]. In particular, the arguments supporting the same conclusions can be united in a single argument.

In any case, many of the computational issues discussed in this and the next section depend only on the semantics and the player's aim, and so are still applicable to these less-well-studied forms of strategic argumentation.

[10] For example, see Proposition 2 of [90].

Finally, even when addressing the same semantics and critical argument, there is some freedom in the strategic aims of the two players. At one extreme the players might have the same aim and, on the other extreme, have diametrically opposed aims. In between these extremes the players might have different but compatible aims, or have incompatible aims. Aims are discussed in detail in the next subsection. In this chapter we assume that the two players have incompatible aims: it is not possible for both players to achieve their aims simultaneously.

In previous sections we have discussed both symmetric and asymmetric forms of strategic argumentation. In abstract argumentation there is no explicit notion of conclusion and, therefore, no notion of an argument supporting the negation of the conclusion of another. Consequently, symmetric strategic argumentation is not available, in general. We will focus on asymmetric strategic argumentation. That is, whatever Pr's aim is, Op's aim is to prevent it.

In summary, a *strategic abstract argumentation dialogue game* consists of a split abstract argumentation framework, a critical argument, an abstract argumentation semantics, and aims for both Pr and Op. The *play* of the game is a sequence of moves such that each player leaves the game in a state where her strategic aim is satisfied.

6.2 Players' Aims

The range of strategic aims a player might have is limited under the grounded semantics. But once we consider semantics with multiple extensions a player has a much wider range.

Initially, work on abstract argumentation focussed on credulous and skeptical acceptance. An argument a in argumentation framework AF under semantics σ is *credulously accepted* if it is labelled `in` in at least one σ-extension. a is *skeptically accepted* if it is labelled `in` in every σ-extension. These two statuses were inherited from the field of non-monotonic reasoning.

[142] extended this work with the notion of justification status. The justification status of an argument a in an argument framework AF is the set of labels a receives in complete extensions. Thus a justification status is a subset of $\{\texttt{in}, \texttt{out}, \texttt{undec}\}$. In general this might lead to $2^3 = 8$ different statuses, but only 6 are possible for the complete semantics [142]. Obviously, this approach can be extended to any extension-based semantics [44].

[91; 90] further extended the range of argument statuses to the following, casting these as possible aims of a proponent

1. **Existential:** a is labelled `in` in at least one σ-extension
2. **Universal:** a is labelled `in` in all σ-extensions

3. **Unrejected:** a is not labelled `out` in any σ-extension
4. **Uncontested:** a is labelled `in` in at least one σ-extension and is not labelled `out` in any σ-extension
5. **Plurality:** a is labelled `in` in more σ-extensions than it is labelled `out`
6. **Majority:** a is labelled `in` in more σ-extensions than it is not labelled `in`
7. **Supermajority:** a is labelled `in` in at least twice as many σ-extensions than it is not labelled `in`

The last three are called *counting aims*, distinct from the first four which are based on zero/non-zero number of labels, like the justification statuses[11]. In addition, the negation of such conditions and their dual (exchanging the role of `in` and `out`), which are plausible aims for the opponent, have also been considered [90].

But clearly there are many more possibilities. Each of the first four strategic aims can be formulated as a disjunction of justification statuses. So we might consider any disjunction of justification statuses as a potential strategic aim. This would give us $2^8 = 256$ strategic aims. Many of these will be unrealizable under some semantics and/or unrealistic in practice. Under the stable semantics, aims that the argumentation framework has at least one extension or has no extension are also sensible. Further possibilities are aims such as: a is accepted in at least 2 extensions or is universally accepted. There are also many variations possible for the counting aims. For example, [91] contemplates a weighting on all extensions, with the arguer's aim that the sum of the weights of extensions in which a is labelled `in` is greater than the sum of weights of the remaining extensions.

Some of the aims seem similar to the ideas behind proof standards that are formalized in [60], although those proof standards are formalized in a very different setting. The Existential aim is similar to a *scintilla of evidence*, the Majority and Supermajority correspond to *preponderance of the evidence* and *clear and convincing evidence*, respectively, while the Uncontested aim is like *beyond a reasonable doubt.*[12] The Universal aim corresponds to *beyond a doubt*, in the phrasing of [51].

There are some obvious close relationships between these different concepts. a is skeptically accepted iff a has justification status {`in`} iff a satisfies the Universal aim. Similarly, a is credulously accepted iff a's justification status contains `in` iff

[11] A counting utility function was defined in [128], but it counts the number of desired conclusions that appear in all σ-extensions rather than counting the number of σ-extensions in which a conclusion appears.

[12] The Uncontested aim is also similar to the notion of *argumentative inference* in paraconsistent reasoning from maximally consistent sets [20].

	$\mathcal{GR}$	$\mathcal{ST}$	$\mathcal{CO}$	$\mathcal{PR}$	$\mathcal{SST}$	$\mathcal{EA}$	$\mathcal{ID}$
Existential	in P	NP-c	NP-c	NP-c	Σ_2^p-c	Π_2^p-c	in Θ_2^p
Universal	in P	coNP-c	in P	Π_2^p-c	Π_2^p-c	Π_2^p-c	in Θ_2^p
Unrejected	in P	coNP-c	coNP-c	coNP-c	Π_2^p-c	Σ_2^p-c	in Θ_2^p
Uncontested	in P	coNP-c	D^p-c	D^p-c	D_2^p-c	Π_2^p-c	in Θ_2^p
Plurality	in P	PP-c	PP-c	in PP^{NP}	in PP^{NP}	Π_2^p-c	in Θ_2^p
Majority	in P	PP-c	PP-c	in PP^{NP}	in PP^{NP}	Π_2^p-c	in Θ_2^p
Supermajority	in P	PP-c	PP-c	in PP^{NP}	in PP^{NP}	Π_2^p-c	in Θ_2^p

Table 1: Complexity of Aim Verification problem for selected strategic aims and semantics [90]. For a complexity class $\mathcal{C}$, $\mathcal{C}$-c denotes that the problem is complete for $\mathcal{C}$.

a satisfies the Existential aim. a satisfies the Unrejected aim iff a has justification status $\{\texttt{in}\}$, $\{\texttt{undec}\}$ $\{\texttt{in}, \texttt{undec}\}$, or $\emptyset$. a satisfies the Uncontested aim iff a has justification status $\{\texttt{in}\}$ or $\{\texttt{in}, \texttt{undec}\}$. Also, a satisfies the Uncontested aim iff a satisfies the Existential and Unrejected aims.

Furthermore, when a semantics consists of a single extension (in particular, the grounded semantics) credulous and skeptical acceptance are identical, there are only three possible justification statuses for an argument ($\{\texttt{in}\}$, $\{\texttt{undec}\}$, and $\{\texttt{out}\}$), and all but the Unrejected aim, of those listed, are identical. In summary, a unitary semantics greatly simplifies analysis of player aims.

Thus, as we consider a wider range of semantics we must also address a wider range of player aims.

6.3 Computational Problems

We can break down the play of a game into two computational problems: recognising whether (or not) an argumentation framework satisfies a given aim, which is called the Aim Verification problem, and determining what arguments to play in order to leave the game in a state where the given aim is satisfied, the decision form of which is called the Desired Outcome problem. These problems will be different for the different players, because they have different aims.

The problem of verifying that an aim is satisfied by some state of strategic argumentation is a fundamental part of each move in a game.

The Aim Verification Problem

Instance A split argumentation framework $(\mathcal{A}_{Com}, \mathcal{A}_{\mathsf{Pr}}, \mathcal{A}_{\mathsf{Op}}, \gg)$, an argumenta-

tion semantics, a critical argument $a \in \mathcal{A}_{Com}$, and an aim.

Question Is the aim concerning the critical argument satisfied under the given semantics by the argumentation framework $(\mathcal{A}_{Com}, \gg)$?

The complexity of this problem, for a selection of semantics and aims, is presented in Table 1. Given Pr's aim, the complexity of verifying Op's aim is the complement of the complexity of Pr's aim.

These results are derived from existing work on the complexity of credulous and skeptical acceptance in abstract argumentation frameworks for the various semantics (see, for example, [43; 141]), and relations between the different aims (Proposition 3 of [90]). For example, the Uncontested aim is the conjunction of Existential and Unrejected, where the latter is the dual of the negation of Existential. Under the (say) preferred semantics, credulous acceptance is NP-complete. Thus the complexity of Uncontested is a conjunction of NP and coNP, which gives us D^p. Completeness is a straightforward reduction.

For the counting aims, clearly the complexity is in PP^V, where V is the complexity of verifying that a set of arguments forms an extension of the appropriate type[13]. The lower bound for the stable semantics is obtained by reduction from the MAJSAT problem, and the complete semantics is treated by reduction from the stable semantics.

Table 1 only addresses a selected set of strategic aims. When a player has such an aim, their opponent will usually have a quite different aim, one not mentioned in the table. Since we are considering only games where the opponent's aim is the complement of the proponent's aim, the complexity of the Aim Verification problem for Op is the complement of the complexity of the Aim Verification problem for Pr. Thus, for example, under the complete semantics, if Pr has the Existential aim then aim verification for Pr is NP-complete, and aim verification for Op is coNP-complete. In general, though, when the opponent's aim is not the complement of the proponent's, the complexity of the two problems is not so directly related.

The Desired Outcome problem [91] is the problem that a player must solve at each step of a strategic abstract argumentation game. It involves identifying that the player has a legal move, leaving the state of the game in a desired state.

The Desired Outcome Problem for Pr

Instance A split argumentation framework $(\mathcal{A}_{Com}, \mathcal{A}_{\mathsf{Pr}}, \mathcal{A}_{\mathsf{Op}}, \gg)$ an argumenta-

[13] There has been some work done on counting extensions, both on the complexity of counting and identifying tractable cases [14; 53]. These works focus on absolute counting, rather than comparing counts (as in the counting aims), so the results are presented in terms of #P rather than PP. Nevertheless, the complexity results are comparable to those for the counting aims in the Aim Verification problem.

	$\mathcal{GR}$	$\mathcal{ST}$	$\mathcal{CO}$	$\mathcal{PR}$	$\mathcal{SST}$	$\mathcal{EA}$	$\mathcal{ID}$
Existential	NP-c	NP-c	NP-c	NP-c	Σ_2^p-c	Σ_3^p-c	Σ_2^p-c
Universal	NP-c	Σ_2^p-c	NP-c	Σ_3^p-c	Σ_3^p-c	Σ_3^p-c	Σ_2^p-c
Unrejected	NP-c	Σ_2^p-c	Σ_2^p-c	Σ_2^p-c	Σ_3^p-c	Σ_2^p-c	Σ_2^p-c
Uncontested	NP-c	Σ_2^p-c	Σ_2^p-c	Σ_2^p-c	Σ_3^p-c	Σ_3^p-c	Σ_2^p-c
Plurality	NP-c	NP^{PP}-c	NP^{PP}-c	NP^{PP}-c	NP^{PP}-c	Σ_3^p-c	Σ_2^p-c
Majority	NP-c	NP^{PP}-c	NP^{PP}-c	NP^{PP}-c	NP^{PP}-c	Σ_3^p-c	Σ_2^p-c
Supermajority	NP-c	NP^{PP}-c	NP^{PP}-c	NP^{PP}-c	NP^{PP}-c	Σ_3^p-c	Σ_2^p-c

Table 2: Complexity of the Desired Outcome problem for Pr, for selected aims and semantics [91; 90; 89]. For a complexity class $\mathcal{C}$, $\mathcal{C}$-c denotes that the problem is complete for $\mathcal{C}$.

tion semantics, a critical argument $a \in \mathcal{A}_{Com}$, and an aim for Pr.

Question Is there a set $I \subseteq \mathcal{A}_{\mathsf{Pr}}$ such that Pr's aim with respect to the critical argument is achieved in the argumentation framework $(\mathcal{A}_{Com} \cup I, \gg)$?

This problem is a generalization of the strategic argumentation problem, as defined in Section 2, which is restricted to accepting the critical argument under the grounded semantics.

It is not difficult to see that the Desired Outcome problem can be solved by a non-deterministic algorithm with an oracle for the Aim Verification problem with Pr's aim. The complexity of this problem, for a selection of semantics and aims, is presented in Table 2.

The complement of this problem decides when Pr does not have a next move. The complexity of this complement problem is clearly the complement of the complexity of the Desired Outcome problem.

We can define the Desired Outcome problem for Op similarly, based on Op's aim. The complexities of the Desired Outcome problems for Pr and Op are not as directly related as is the case for aim verification.

Showing the presence of the Desired Outcome problem in the appropriate complexity class is comparatively straightforward, but showing it is complete in the class requires the construction of argumentation frameworks that extend those used for credulous or skeptical acceptance. An example construction for the Desired Outcome problem with the Universal aim under the stable semantics is shown in Figure 2. In this case the problem is Σ_2^p-complete, so we reduce the satisfiability of $\exists X \forall Y\ \psi$ (where ψ is in DNF) formulas to this problem. The diagram has three parts: on the

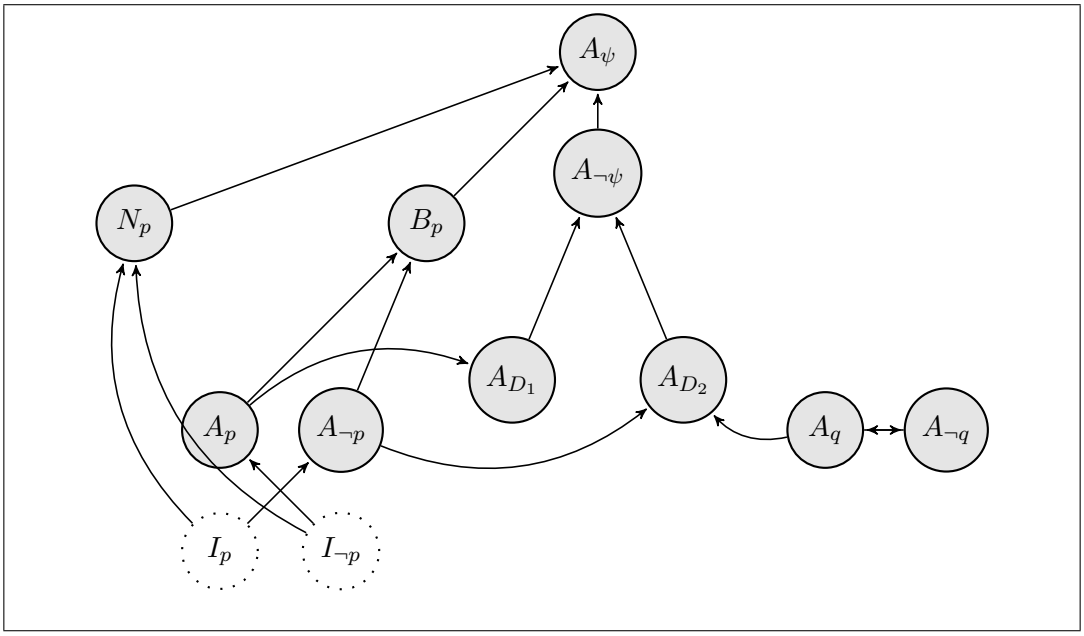

Figure 2: Example construction for the Desired Outcome problem with the Universal aim under the stable semantics

left is the representation of a variable p in X, in the middle is the representation of ψ, and on the right is the representation of a variable q in Y.

In the diagram, the grey nodes are arguments in $\mathcal{A}_{Com}$, and the white nodes (I_p and $I_{\neg p}$) are arguments in $\mathcal{A}_{\mathsf{Pr}}$. $\gg$ is described by the directed edges. ($\mathcal{A}_{\mathsf{Op}}$ is irrelevant to this problem.) Intuitively, an argument A_s (where s is a literal) accepted in a stable extension corresponds to the literal s being true. The critical argument is A_ψ, and Pr must move so that A_ψ is accepted in all stable extensions. The construction ensures that if Pr plays either both I_p and $I_{\neg p}$ or neither I_p nor $I_{\neg p}$ then either B_p or N_p is accepted and A_ψ is rejected in all stable extensions. Thus, Pr must play only one argument for each p, and this ensures only one of A_p and $A_{\neg p}$ can be accepted. This part of the construction is common to all reductions.

In the diagram, the formula is $\exists p \forall q\ \neg p \vee (p \wedge \neg q)$. It is represented in a slightly roundabout way. The treatment of variables q in Y ensures that both stable extensions containing A_q (i.e. q is true) and stable extensions containing $A_{\neg q}$ (i.e. q is false) are generated. A more formal description of this construction is in the proof of Theorem 7 of [91].

Given a specific game, we write AV_{Pr} (AV_{Op}) for the Aim Verification problem for Pr's (respectively, Op's) aim. Similarly, DO_{Pr} (DO_{Op}) denotes the Desired Outcome

problem for Pr (respectively, Op).

Play begins by Pr playing a set of arguments, including the critical argument, and proceeds by Op and Pr alternately solving their Desired Outcome problem and playing the corresponding set of arguments. Play can extend for, at most, $\min(|\mathcal{A}_{\mathsf{Pr}}|, |\mathcal{A}_{\mathsf{Op}}|)$ rounds before play terminates, when one player does not make a move. Thus, play for Pr, over the entire game, has a computational cost in $\mathrm{P}^{DO_{\mathsf{Pr}}}$ while the cost of play for Op is in $\mathrm{P}^{DO_{\mathsf{Op}}}$ [90].

The Aim Verification problem is of little interest for the concrete forms of strategic argumentation discussed in Section 5. In those cases the inference problem is polynomial [83; 23]. Consequently, verifying any of the aims or justification statuses is also polynomial. The Desired Outcome problem corresponds to the SSA, SSSA and AsSA problems in Section 5: they represent the computational cost of making a move, in their respective games. In the case of structured arguments, conceptually the argumentation theory gives rise to an argumentation framework, which can then be interpreted in a chosen semantics. However, this does not mean that the NP-completeness for grounded semantics in Table 2 can be used to prove Theorem 5.8. The difficulty is that there might be greater than polynomially many arguments generated from the argument theory.

7 Corruption in Argumentation

When a game such as strategic argumentation is a model of a real-world situation, we must acknowledge the extra forces and influences that operate upon a player, beyond those of the specific role they have in the game. Often a player is assumed to have no motivations beyond performing their role and conforming to the rules of the game, but this is a rather simplistic view. While games do have rules, we need to consider the possibility that a player breaks the rules, or "cheats".

The context of the game is important in this regard. Organizations have many mechanisms to discourage the risk of corruption of their processes by the individuals performing these processes: managerial oversight, transparency through audit trails, the presence of co-workers, random inspections, etc. Society, as a whole, provides an entire justice system to enforce the rules the society considers important, and to detect and punish violations. When these mechanisms are not available, or are limited, how can we discourage rule-breaking?

[16] proposed an answer to this question in the case of vote manipulation: if the computational difficulty of determining what an individual must do to alter the result of an election is too great, a potential vote manipulator may be discouraged from the manipulation, even though he has the opportunity to do it. They called this

concept *computational resistance to strategic manipulation*. This insight has spawned a whole subfield of computational social choice [29]. In this section we describe the application of these ideas to strategic argumentation.

Throughout this section, we consider that players are engaged by a client to play the game. A player is expected to adhere to the rules of the game and, in particular, play the game to win for her client. However, while the client is invested in winning the game, the player has various competing incentives. These are the source of the corruption we consider. A player might cheat on behalf of her client, or might sacrifice her client's chances for other incentives. This issue is known in management theory as the *principal-agent problem* or the *agency problem* [47].

7.1 Corruption and Resistance

Strategic argumentation has relatively few rules, though some are implicit rather than explicitly stated. The players must take turns, but violations of this rule are obvious and, anyway, offer no advantage to the players. A player must make a move if one is available to her. This rule is implicit in the assumption that the player will play her role properly. Such a rule is difficult to enforce without knowledge of the player's arguments. The player's arguments are assumed to be private, but this is also difficult to enforce. We will focus on violations of this privacy[14].

We consider two forms of corruption. The first, *collusion*, arises when one player induces the other to let her win. Such behaviour on its own is straightforward, though illicit, and does not, as such, appear in the game. But it is complicated by the desire of the guilty parties not to be detected. Thus, colluding players must not only ensure the "right" player wins, they must also make sure that an external observer cannot distinguish the collusive play from normal play. If the work to ensure this is computationally more difficult than simply playing the game honestly, then we consider the game to be *resistant to collusion*.

The following example is an instance of collusion.

Example 7.1. *Consider the strategic argument game depicted in Figure 3, where vertices are arguments (grey if they can be played by* Pr*, white for* Op*) and edges are attacks of one argument on another. For concreteness, we assume that we employ the grounded semantics and* Pr*'s strategic aim is that argument* A *is accepted. Normal play would proceed as follows:* Pr *plays* A*,* Op *plays* B_1 *(thus defeating* A*),* Pr *plays* C *(restoring* A *by defeating* B_1*), and* Op *plays* D *(defeating* C*, and allowing* B_1 *to defeat* A*). Thus, normally,* Pr *loses.*

[14] Earlier works that consider privacy include [32] and [105], which have a focus on minimizing the exposure of a player's arguments during play, rather than the loss of privacy by corruption.

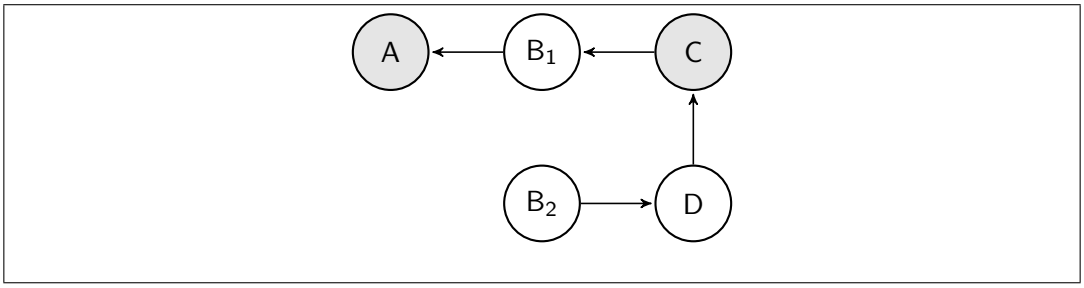

Figure 3: A strategic argumentation game. An argument is grey if it can be played by Pr and white if it can be played by Op.

However, Pr *and* Op *might collude to ensure* Pr *wins by playing as follows:* Pr *plays* A*,* Op *plays* B_1 and B_2*, and* Pr *plays* C *(restoring* A*).* Pr *now wins because* Op *has no effective move: to play* D *would have no effect because it is defeated by* B_1*.*

This example also serves to show the difference between collusion and an omniscient argumentation framework $(\mathcal{A}_{Com} \cup \mathcal{A}_{\mathsf{Pr}} \cup \mathcal{A}_{\mathsf{Op}}, \gg)$*. Under any completist semantics,* A *is accepted in the omniscient argumentation framework, but if* Pr *and* Op *collude to ensure* Op *wins they can do so by following the normal play above.*

The second form of corruption, *espionage*, occurs when, through some means, one player gains knowledge of the other player's arguments. Again, this act is not apparent in the game, but it requires work to develop a strategy, based on that knowledge, to defeat the other player. If this is computationally more difficult than playing the game honestly, then we consider the game to be *resistant to espionage*.

In Example 7.1, the corrupt sequence of moves might also occur if Op committed espionage on Pr in order to ensure Pr wins.

For both forms of resistance, we need to clarify what "computationally more difficult" means. Computational difficulty will be measured in terms of a hierarchy of complexity classes where, although one class might be contained in another, it is often not known that the two classes are distinct. However, if the two classes were equal then part of the (say) polynomial complexity hierarchy would collapse, and this is commonly believed by complexity theorists not to happen. Thus "computationally more difficult" is subject to this commonly-believed assumption. For counting aims we are dealing with the counting polynomial hierarchy, and the corresponding assumption is messier. The topic is less investigated and there are some collapses known within the counting hierarchy. However, those collapses do not affect the containments

$$\mathrm{P}^{\mathrm{PP}} \subseteq \mathrm{NP}^{\mathrm{PP}} \subseteq \mathrm{P}^{\mathrm{NP}^{\mathrm{PP}}} \subseteq \mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}} \subseteq \cdots \subseteq \mathrm{PSPACE}$$

The assumption that these containments are strict is the basis of resistance for counting aims.

Inherent in the resistance approach to corruption is the assumption that players will be effectively penalised if their corruption is detected. This assumption relies on issues of governance, lasting identification of the players, and enforcement and scale of penalties, among others. But these issues depend on the context of the game and are beyond the scope of this chapter.

7.2 Computational Problems

We now consider the computational problems that must be solved by players in order to exploit corruption.

Colluders need to to construct an alternating sequence of moves that ends with Pr winning, that is, with Op unable to make a move. This is formalized as follows.

The Winning Sequence Problem for Pr

Instance A split argumentation framework $(\mathcal{A}_{Com}, \mathcal{A}_{\mathsf{Pr}}, \mathcal{A}_{\mathsf{Op}}, \gg)$ and a desired outcome for Pr.

Question Is there a sequence of moves such that Pr wins?

A similar problem arises when the colluders wish to ensure that Op wins.

The problem for Pr can be solved by nondeterministically generating a sequence of moves, verifying that each move achieves the aim for its player, and verifying that Op has no further move. That is, it can be solved in NP with oracles for AV_{Pr}, AV_{Op} and (the complement of) DO_{Op}. $AV_{\mathsf{Op}} = coAV_{\mathsf{Pr}}$, since we assume Pr and Op have complementary aims, so the larger of $\mathrm{NP}^{AV_{\mathsf{Pr}}}$ and $\mathrm{NP}^{DO_{\mathsf{Op}}}$ is an upper bound for this problem.

In the case of espionage, one player, say Pr, knows her opponent's arguments $\mathcal{A}_{\mathsf{Op}}$ and desires a strategy that will ensure Pr wins, no matter what moves Op makes. A *strategy* for Pr in a split argumentation framework $(\mathcal{A}_{Com}, \mathcal{A}_{\mathsf{Pr}}, \mathcal{A}_{\mathsf{Op}}, \gg)$ is a function from a set of common arguments to the set of arguments to be played in the next move. A sequence of moves $S_1, T_1, S_2, T_2, \ldots$ resulting in common arguments $\mathcal{A}_{Com}^{\mathsf{Pr},1}, \mathcal{A}_{Com}^{\mathsf{Op},1}, \mathcal{A}_{Com}^{\mathsf{Pr},2}, \mathcal{A}_{Com}^{\mathsf{Op},2}, \ldots$ is *consistent with* a strategy s for Pr if, for every j, $S_{j+1} = s(\mathcal{A}_{Com}^{\mathsf{Op},j}, \mathcal{A}_{\mathsf{Pr}})$. A strategy for Pr is *winning* if every valid sequence of moves consistent with the strategy is won by Pr.

The Winning Strategy Problem for Pr

Instance A split argumentation framework $(\mathcal{A}_{Com}, \mathcal{A}_{\mathsf{Pr}}, \mathcal{A}_{\mathsf{Op}}, \gg)$ and a desired outcome for Pr.

	$\mathcal{GR}$	$\mathcal{ST}$	$\mathcal{CO}$	$\mathcal{PR}$	$\mathcal{SST}$	$\mathcal{EA}$	$\mathcal{ID}$
Existential	Σ_2^p-c	Σ_3^p-c	Σ_3^p-c	Σ_3^p-c	Σ_4^p-c	Σ_3^p-c	Σ_3^p-c
Universal	Σ_2^p-c	Σ_2^p-c	Σ_2^p-c	Σ_3^p-c	Σ_3^p-c	Σ_3^p-c	Σ_3^p-c
Unrejected	Σ_2^p-c	Σ_2^p-c	Σ_2^p-c	Σ_2^p-c	Σ_3^p-c	Σ_4^p-c	Σ_3^p-c
Uncontested	Σ_2^p-c	Σ_2^p-c	Σ_3^p-c	Σ_3^p-c	Σ_4^p-c	Σ_3^p-c	Σ_3^p-c
Plurality	Σ_2^p-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	Σ_3^p-c	Σ_3^p-c
Majority	Σ_2^p-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	Σ_3^p-c	Σ_3^p-c
Supermajority	Σ_2^p-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	$\mathrm{NP}^{\mathrm{NP}^{\mathrm{PP}}}$-c	Σ_3^p-c	Σ_3^p-c

Table 3: Complexity of the Winning Sequence problem for Pr for selected aims and semantics [90].

Question Is there a winning strategy for Pr that satisfies the standards?

There is also, of course, the corresponding problem for Op which arises when Op conducts the espionage.

The following result shows that the Winning Strategy problem is PSPACE-complete for all completist semantics and all the aims discussed in this chapter. This is not surprising since, as a result of the espionage, Pr is essentially playing a complete knowledge game and such games are known to be PSPACE-hard, in general.

Theorem 7.2. *[90] Consider any completist semantics for abstract argumentation, and any of the above aims for* Pr*.*

The Winning Strategy problem is PSPACE-complete.

This theorem applies both to espionage by Pr and espionage by Op. The constructed argumentation framework for this proof is well-founded. Consequently the construction serves for all completist semantics.

7.3 Audit: Standards and Compliance

To investigate collusion, we need to understand what "normal play" looks like and how to recognise it. [92] proposes that we view this as a matter of audit, with an external body setting standards for play and testing for compliance. In this view there can be multiple standards. We have already seen one standard: that a player must make a move, if she has one (we will call this the *compulsory move* standard). Consequently, colluding players must arrange their play to ensure that the designated loser has no possible moves at the end of the game. Earlier work [91; 90; 89] implicitly operated under this standard.

However, this standard fails to address obvious collusion, like that in Example 7.1. Thus, additional standards are required. However, a standard can only be justified if it does not interfere with honest play. That is, a player should never face a choice between following the standard and improving her chances of winning. Otherwise, any violation of the standard can be explained away as an attempt to improve those chances.

It is clear that the problem in Example 7.1 stems from Op playing B2. But it is not clear what is an appropriate standard that would prevent this move. Several possibilities suggest themselves:

(1) A player should not play an argument that attacks one of her own (unplayed) arguments, thus causing a self-inflicted injury.

(2) A player should play the smallest number of arguments to achieve her aim[15].

(3) A player should play a subset-minimal set of arguments that achieve her aim.

(1) is clearly too strong to be a standard. If, in Example 7.1 (Figure 3), B_1 also attacked B_2 then following this standard would cause Op to lose immediately. However, when the omniscient argumentation framework is known to a player, [128] prove that this standard (which they call the overcautious selection function) is dominant. Unfortunately, a player cannot be expected to know the omniscient argumentation framework.

(2) is more plausible, but consider the following example from [92].

Example 7.3. *Consider the strategic argumentation game in Figure 4, and play that proceeds as follows:* Pr *plays* A, Op *plays* B_1 *and* B_2, *and* Pr *plays* C_1 *and* C_2. *At this stage O must attack both* C_1 *and* C_2, *and she has two alternatives: (1) play* E, *which attacks both* C_1 *and* C_2, *or (2) play both* D_1 *and* D_2, *each attacking one of the* C *arguments. Clearly (1) is the minimum cardinality move. However,* Pr *then responds with* F, *and wins. In (2), the play of* F *is insufficient for* Pr, *since* B_2 *remains undefeated. Hence* Op *wins.*

Thus minimum cardinality is not suitable as a standard, because it can prevent a winning move.

However, [92] showed that (3) is compatible with normal play: every non-minimal move is dominated by a minimal move[16]. Thus the requirement to play only subset-minimal moves is a suitable standard. It remains open whether there are other standards that could be applied.

[15] This is similar to the heuristic of [105], though the details of the game are different.

[16] Previous work addressing redundancy or relevance in argumentation includes [54; 98].

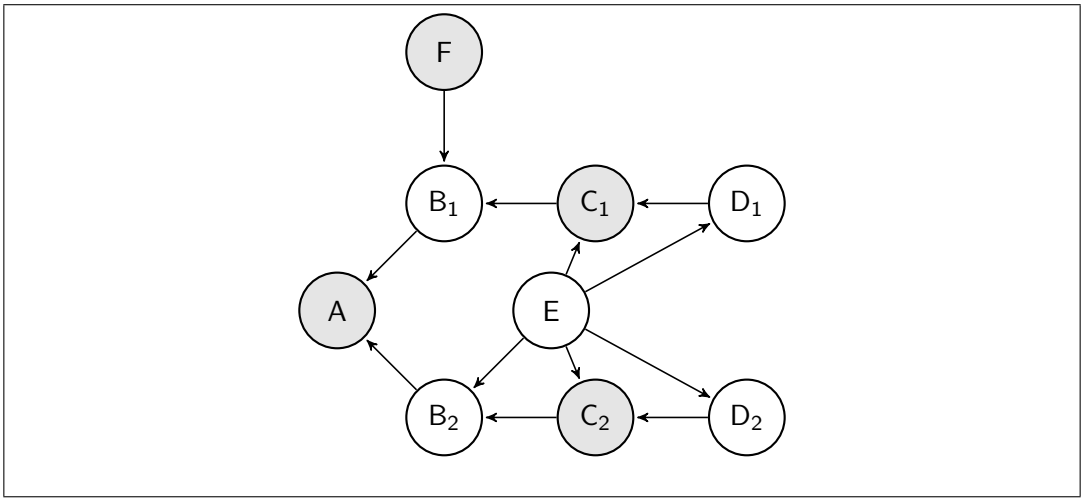

Figure 4: Split argumentation framework demonstrating non-dominance of minimum cardinality moves.

In addition, we need to consider how play can be verified as compliant with a standard. This involves issues of which data need to be accessed by the auditor, as well as the computational difficulty of verifying compliance

In terms of accessibility, all that an auditor needs for subset-minimality is an ability to inspect the initial $\mathcal{A}_{Com}$, the sequence of moves, and $\gg$ restricted to the current $\mathcal{A}_{Com}$, all of which can be considered public information. On the other hand, to verify the compulsory move standard requires knowledge of the player's arguments, which is private. Thus an auditor verifying both standards needs access to all aspects of a split argumentation framework. (However, each client might be in a position to audit the compulsory move standard, which would allow the player's arguments to be kept private from the auditor.)

For the auditor, the cost of verifying compliance with the subset-minimality standard involves polynomial many solutions of the Minimal Move problem (see next subsection) for Pr, and the same for Op. In comparison, the compulsory move standard requires a $coDO_{\mathsf{L}}$ check, where L is the loser of the game, to verify that there is no move for L left to play.

For the players, compliance with the subset-minimality standard increases the difficulty of making a move. Not only must they find a move, they must also verify that it is minimal. It also increases the cost to players exploiting collusion: they must arrange the game so that their designated player wins, but also ensure that each move is minimal. Furthermore, one easy avenue for exploiting collusion has

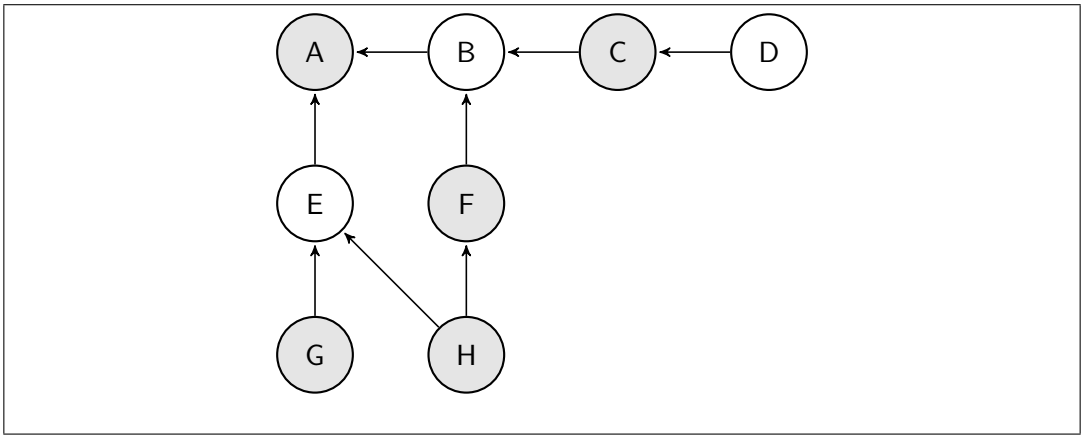

Figure 5: A strategic argumentation game demonstrating weakness of the compulsory move and subset-minimality standards.

been eliminated. Consequently, there are games (like Example 7.1) where compliance with both standards ensures that exploitation of collusion cannot be disguised as normal play.

This leads to some questions. Are these two standards sufficient to prevent the disguise of collusion? If not, can we add standards to achieve this goal? Unfortunately, the answer to the first question is no, as the following example shows.

Example 7.4. *Consider the strategic argument game depicted in Figure 5, where arguments in* $\mathcal{A}_{\mathsf{Pr}}$ *are grey and arguments in* $\mathcal{A}_{\mathsf{Op}}$ *are white, and* A *is the critical argument. If* Pr *refrains from playing* H *then* Pr *will win, since the two arguments attacking* A *(*B *and* E*) can be attacked by* Pr*'s arguments* F *and* G*, which cannot be attacked by* Op*. For example, the sequence of moves:* $\mathsf{A}, \mathsf{B}, \mathsf{F}, \mathsf{E}, \mathsf{G}$ *results in* Pr *winning.*

On the other hand, the sequence of moves: $\mathsf{A}, \mathsf{E}, \mathsf{H}, \mathsf{B}, \mathsf{C}, \mathsf{D}$ *results in* Op *winning. Thus,* Pr *and* Op *can collude to ensure* Op *wins.*

This example suggests that a variation of (1) above might be needed to detect collusion more thoroughly. Which leads us to the second question: is it possible to impose enough justified standards that no collusion can be disguised as compliant play? Again the answer is no.

Consider the argumentation game in Figure 6 under the grounded semantics, where A is the critical argument. After Pr plays A, Op has the choice of playing B or C. Depending on this choice, either Pr or Op will win. If Pr and Op collude they can determine the outcome, but any real restriction imposed by a standard will restrict

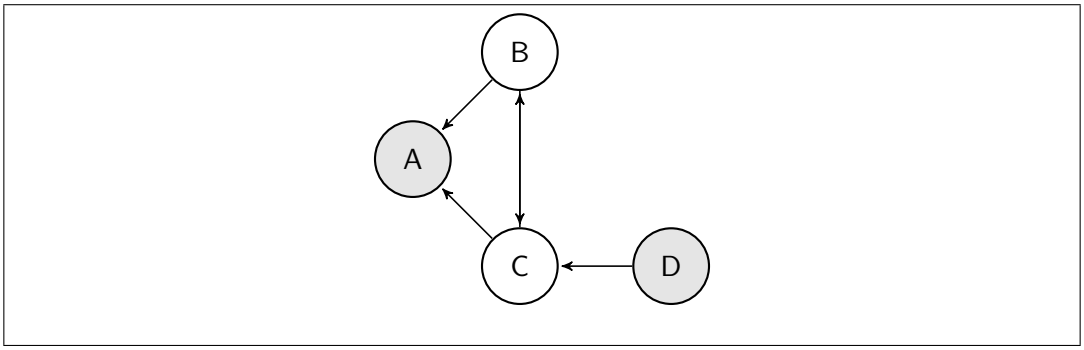

Figure 6: A strategic argumentation game demonstrating that no accumulation of justifiable standards can make all collusion detectable.

to one possible outcome, so it cannot be a justified standard. Thus any collusion in this game cannot be detected by imposing justified standards.

Hence, we see that collusion cannot be prevented simply by imposing more and more standards. We must continue to rely on computational difficulty to discourage corruption.

We now take a stab at formalizing these considerations. A *standard* is a restriction on moves a player may make. More precisely, a standard is a function from a player's aim, her private set of arguments ($\mathcal{A}_{\mathsf{Pr}}$ or $\mathcal{A}_{\mathsf{Op}}$), a proposed move (a subset of her private arguments), and the set of arguments $\mathcal{A}_{Com}$, that are common knowledge, to the set $\{permitted, not\_permitted\}$. The standard is *complied with* by a player in the play of a game if each move by the player is permitted by the standard.

A set of standards is *justified* if, for every argumentation game, if for every unpermitted move that achieves the player's aim there is a better (or equal) permitted move that achieves the player's aim. A move m by a player is considered better or equal to another move m' if, for every behaviour of the opposing player, the player can achieve a better or equal outcome of the game by playing m, rather than playing m'. Note that a set of standards might be unjustified even though each standard, individually, is justified. However the combination of the compulsory move and subset-minimality standards is justified.

We say that a strategic argumentation game played under a given finite set of standards has *detectable collusion* if any occurrence of collusion that affects the outcome of the game violates a standard. The set of standards must be finite because an infinite set of standards creates difficulties for compliance verification, both for the players and the auditor. The best that could be done is checks on a random subset of standards. On the face of it, this might be sufficient for the auditor, but if the player has no way to verify her move is compliant with all standards then the

auditor cannot reliably infer collusion or incompetence from her failure to comply.

We say that a strategic argumentation game played under a given set of standards is *determined* if all compliant plays of the game lead to the same winner. It appears that collusion is detectable iff the game is determined.

These considerations are similar to the issues in game-theoretic *mechanism design* (see, for example, [57]) where the aim of the design is to achieve some social good, such as fairness, honesty, ..., despite the self-interest of the parties involved. Thus there is a strong focus is on a strategy-proof mechanism, where there is no advantage to players in deviating from socially good behaviour. A classic example of mechanism design is two-person cake-cutting, where the mechanism specifies that one player cuts the cake in two, and the other chooses a piece. This mechanism encourages fairness in the division of the cake.

In an argumentation setting, [116] addresses a version of strategic abstract argumentation (with multiple players) where all players simultaneously play a selection of their arguments, aiming for their focal argument to be accepted. The social good desired is that the arguments accepted after all moves are those that would be accepted if all arguments were available (the omniscient view of the split argumentation framework). That is, roughly, the social good is that arguments are not hidden[17]. Other work, such as [126; 122], also considers hiding of arguments as unfair or dishonest.

This is a different attitude than in strategic argumentation, which treats argument hiding as an inherent feature of adversarial argumentation. [116] characterize when their game is strategy-proof, that is, when there is no advantage to players from hiding arguments. It is only in very restrictive circumstances that honesty is the best policy. Their focus is on the game itself. In particular, the self-interest players have derives from their goals within the game. This is in common with most work on mechanism design. In contrast, the work in this section aims at aligning the self-interest of players with their clients, where that self-interest extends *beyond the game itself*. The introduction of standards is an instance of mechanism design, but we have seen that there is no mechanism that allows strategic play and prevents all collusion. Consequently, computational resistance serves as a back-stop, to discourage collusion.

[17] An argument a is *hidden* if it is not played, even though a player has it available to play. Sometimes, more specifically, it refers to an argument a that defeats an argument b, but is not played when b is played.

	$\mathcal{GR}$	$\mathcal{ST}$	$\mathcal{CO}$	$\mathcal{PR}$	$\mathcal{SST}$	$\mathcal{EA}$	$\mathcal{ID}$
Existential	*Res*	*Res*	*Res*	*Res*	*Res*		*Res*
Universal	*Res*		*Res*				*Res*
Unrejected	*Res*					*Res*	*Res*
Uncontested	*Res*		*Res*	*Res*	*Res*		*Res*
Plurality	*Res*	*Res*	*Res*	*Res*	*Res*		*Res*
Majority	*Res*	*Res*	*Res*	*Res*	*Res*		*Res*
Supermajority	*Res*	*Res*	*Res*	*Res*	*Res*		*Res*

Table 4: Resistance to collusion to ensure Pr wins, for several aims and semantics [90]. *Res* denotes that the combination of aim and semantics is computationally resistant to collusion, while a blank denotes that it is not resistant.

7.4 Resistance to Corruption

Recall that resistance to collusion is based upon the relative computational difficulty of exploiting the corruption, while disguising it, versus the difficulty of playing the game honestly. In other words, we compare the complexity of the Winning Sequence problem with the complexity of normal play as described at the end of subsection 6.3. This comparison is presented in Table 4. While not all combinations of aim and semantics show computational resistance, many do. However, it is notable that three of the aims under the stable semantics do not have resistance to collusion.

This comparison, however, deals only with the initial standard: that a player must play if she has a move. We need to recalculate both the computational cost of normal, honest play and the complexity of the Winning Sequence problem under both standards, in order to determine resistance to collusion when both standards apply. Hence, we need to consider the computational cost of verifying compliance with the subset-minimality standard. The Minimal Move Problem is to verify that a given move is a subset-minimal move.

The Minimal Move Problem for Pr

Instance A split argumentation framework $(\mathcal{A}_{Com}, \mathcal{A}_{\mathsf{Pr}}, \mathcal{A}_{\mathsf{Op}}, \gg)$, an argumentation semantics, an aim for Pr, and a move $M \subseteq \mathcal{A}_{\mathsf{Pr}}$ that achieves the aim for Pr.

Question Is M a minimal set that achieves the aim under the given semantics? That is, is there no subset $N \subset M$ such that Pr's desired outcome is achieved in the

argumentation framework $(\mathcal{A}_{Com} \cup N, \gg)$?

It is clear that the complement of this problem can be solved by a non-deterministic algorithm that guesses N and uses an oracle for the Aim Verification problem. Thus the Minimal Move Problem is in coNP^{AV}, where AV is the complexity of the Aim Verification problem. The complexity of the Minimal Move problem for Pr and Op (denoted by MM_{Pr} and MM_{Op}) for selected aims (of Pr) and the grounded and stable semantics is given in Table 5. This is also the work that an auditor must do to verify compliance with the subset-minimality standard. All aims for the grounded semantics lead to the same complexity, so these results have been condensed to a single row.

Honest (i.e. non-corrupt) play under both standards consists of a polynomial number of moves, each involving the search for an effective move, incorporating a verification that the aim is satisfied and the move is minimal. The cost of a single move for Pr is DOM_{Pr}, which is in $\text{NP}^{\{AV_{\mathsf{Pr}}, MM_{\mathsf{Pr}}\}}$ and the total cost of honest play is $\text{P}^{DOM_{\mathsf{Pr}}}$, and similarly for Op. The total cost of honest play for each player, under the two standards, is shown in Table 5. In some cases the complexity of play has increased as a result of the additional standard, but in other cases it remains the same.

Finally, we must recalculate the cost for collusive play (assuming the players want Pr to win), denoted by WSM. This is the cost of solving the Winning Sequence problem when each player is constrained by the standard to play only subset-minimal moves. The players must search for a sequence of effective minimal moves, and ensure Op has no effective move remaining. Thus WSM is in $\text{NP}^{\{AV_{\mathsf{Pr}}, MM_{\mathsf{Pr}}, AV_{\mathsf{Op}}, MM_{\mathsf{Op}}, coDO_{\mathsf{Op}}\}}$. The complexity of WSM is also given in Table 5. In most cases the additional standard does not change the complexity of solving the Winning Sequence problem.

We can see from the table that, once the subset-minimality standard is incorporated, all aims under the stable semantics are resistant to collusion, an improvement (compare with Table 4).

While the additional standard may increase the cost of playing a strategic argumentation game, it is still not comparable to the cost of solving the Winning Strategy problem. Hence all the completist semantics and all the aims remain resistant to espionage.

Of all the semantics that have been investigated, the naive semantics has an interesting property – it is *corruption-proof*, at least for the non-counting aims [89]. Under this semantics the extensions are the maximal conflict-free sets. It is corruption-proof because the outcome is determined by the arguments the players have, if they comply with the compulsory move standard. In this sense, the game is

	MM_{Pr}	MM_{Op}	Hon^{Min}_{Pr}	Hon^{Min}_{Op}	WSM	
Grounded semantics	coNP-c	coNP-c	Δ^p_2-c	Δ^p_2-c	Σ^p_2-c	*Res*
Stable semantics						
Existential	coNP-c	Π^p_2-c	Δ^p_2-c	Δ^p_3-c	Σ^p_3-c	*Res*
Universal	Π^p_2-c	coNP-c	Δ^p_3-c	Δ^p_2-c	Σ^p_3-c	*Res*
Unrejected	Π^p_2-c	coNP-c	Δ^p_3-c	Δ^p_2-c	Σ^p_3-c	*Res*
Uncontested	Π^p_2-c	coNP-c	Δ^p_3-c	Δ^p_2-c	Σ^p_3-c	*Res*
Plurality/Majority	coNP^{PP}-c	coNP^{PP}-c	$\text{P}^{\text{NP}^{\text{PP}}}$-c	$\text{P}^{\text{NP}^{\text{PP}}}$-c	$\text{NP}^{\text{NP}^{\text{PP}}}$-c	*Res*
Supermajority	coNP^{PP}-c	coNP^{PP}-c	$\text{P}^{\text{NP}^{\text{PP}}}$-c	$\text{P}^{\text{NP}^{\text{PP}}}$-c	$\text{NP}^{\text{NP}^{\text{PP}}}$-c	*Res*

Table 5: Complexity of Minimality problems and normal play with the minimality standard (for Pr and Op), Winning Sequence problems (for Pr), and resistance to collusion (to ensure Pr wins), under the grounded and stable semantics, for selected aims (of Pr) [92].

strategy-proof. Consequently, if the game has an outcome different from the expected one, we detect corruption/incompetence. But, since every game is determined, this is not a suitable semantics in which to do strategic argumentation.

7.5 Concrete Argumentation Systems

As we saw in Section 5, the SSA, SSSA, and AsSA problems for DL(∂) and DL(δ) are NP-complete, as are the problems for the ASPIC-like language under the grounded semantics. These correspond to the Desired Outcome problem. It was shown in [88] that the Winning Strategy problem is PSPACE-complete and the Winning Sequence problem is Σ^p_2-complete for DL(∂); hence, argumentation in DL(∂) is resistant to corruption. These results relied on careful constructions and proofs reliant on the specific logic.

There are many concrete languages, beyond those discussed in Section 5, that can be used to express arguments. There is a wide variety of defeasible logics [23; 96; 95; 22; 94], languages incorporating inheritance in logic programming [78; 28], other logic programming-based languages [139; 140; 76; 123], languages inspired by argumentation [38; 135], as well as primitive systems like non-monotonic inheritance networks [129]. Unlike systems such as ASPIC [2; 112; 143] and assumption-based argumentation [26], these languages are designed independently from – and sometimes prior to – abstract argumentation. Thus the results of this section do not apply directly to such languages, and following the approach of [88] to establish resistance

to corruption would be time-consuming.

However, it was shown in [93] that many of these concrete languages can encode abstract argumentation frameworks under appropriate semantics. Most of the languages employ the grounded semantics, while DefLog [135], ASPDA [140] and a version of NDL [95] employ the stable semantics. Similarly, defeasible logics defined in the framework of [5] for a range of logic programming semantics can encode corresponding abstract argumentation frameworks under the corresponding (in the sense of [31]) completist semantics. As a result, the hardness complexity results for these semantics are carried over to the concrete languages. Consequently, many of these languages are shown to be resistant to corruption. See [91] for details.

8 Related Work

Dialogue games for argumentation describe systems where two opponents argue about the tenability of one or more claims (and thus are in the class of persuasion dialogues [138]). Persuasion dialogues are typically substantive: the participants provide substantive reasons for their claims [81]. As a consequence, the information available during the game evolves, each participant discovering new pieces of information each time the opponent makes new claims.

A structural difference between strategic argumentation and many persuasion dialogues lies within the nature of the reply/counter-argument a player may present: in our setting a participant never asks a *why?* question to a previous opponent's claim. In fact, the answer to the *why?* question is already provided at the very moment a claim is made: every and each claim is justified/supported by the argument proving it (all the rules in the proof of which the claim is the conclusion). Dialogue systems have been classified based on their structural properties, that is whether a player can make a single or multiple moves in one turn, and whether she is allowed to reply only once or multiple times to the other player's moves. In our game, the turn shifts immediately after a player's move, but this is nonetheless a relaxed constraint given that, during such a move, the player may advance a set of arguments, and not just a single one. Moreover, the player is not obliged to respond to the opponent's last move but she may attack any argument proposed so far (possibly her own if this can prove her claim). It is nonetheless true that our framework is a a sort of *unique/move* protocol (a hybrid version): a player can respond only once before the turn passes to the other player even if, as we have shown, such a response is not limited to a single argument.

We do not allow argument retractions (also known as withdrawals): once an argument is played, it will remain as part of the common rules/knowledge base till

the end of the game. But it is clear that such a constraint does not prevent a player attacking one of her previously played arguments. We force a replying move to be structurally *relevant*, that is it must be capable of changing the dialogical status of the critical literal/argument (except for the surrendering move which, instead, gives the victory to the adversary). Even allowing retraction in our framework, the computational complexity does not change: a retraction operation would choose a set of rules/arguments to be discarded; thus there is still a choice to be made. However, retraction would change the nature of the game: in the game of Figure 6, Op would not lose. Furthermore, retraction requires restrictions to ensure games terminate.

On the other hand, within our framework a player is not committed to the arguments she plays. Commitments typically require that moves do not contradict or challenge previous commitments/statements; in our framework, players have commitments only towards the claim at dispute as they may, at any time, advance arguments contradicting their own previous statements.

Our turn-taking is in line with the notion of [109; 82] where "when a player is to move, s/he keeps moving until s/he has changed the status of the initial move his or her way". The sole difference is that we consider the playing of more arguments as a single move, but the essential idea is that even in our framework the player must change the status of the initial claim (the critical literal/argument).

The structure of the arguments defined by our framework is in line with [109]. The idea of an argumentation theory is that of containing all the arguments that are constructible on the basis of a certain theory or knowledge base.

Our framework is *sound* and *fair* according to definitions given in [109]. It is sound because if the proponent wins the game, then the current theory actually proves the critical literal. (Symmetrically, if the opponent wins, the theory either fails to prove the critical literal, disproves it, or proves the opposite, depending on the game variant.) The framework also satisfies fairness given that if, at a given turn, the theory proves the critical literal, then proponent is winning the game. (Again, depending on the type of game, we have that if the theory either fails to prove the critical literal, disproves it, or proves the opposite at a given turn, then the opponent is winning the game.)

The conceptual basis of our formalisation that an argument moved at some earlier stage might be a legal counterargument against some later arguments is not a novelty in the literature of the field, and has been adopted in many frameworks [108; 109].

Our dialogues are coherent (in the sense proposed by [109, Section 7.1]) since we do not allow players to retract their claims. A participant can play a set of arguments conflicting with some of the moves she has put forward in previous steps, if this helps her in taking advantage of information disclosed by the adversary.

[36] describe a rigorous persuasion dialogue game $\mathrm{RPD}_{\mathrm{GD}}$ obtained by adapting the game RPD_0 of [138], replacing propositional logic as the underlying information carrier with abstract argumentation. It has some features in common with strategic argumentation, including private arguments, alternating moves and strategic play. On the other hand, each move is a single locution, which may be a statement, challenge, or question; the only semantics considered is the grounded semantics; and the roles of Pr and Op are quite different from each other, in comparison to strategic argumentation. [36] analyse strategies for their game but it is unclear whether they could be adapted to strategic argumentation.

In game-theoretic terms, a player in a strategic argumentation game has *perfect information* of the structure of the game, the history of the game, and the effects of each move. On the other hand, the players have *incomplete information* of the arguments – and, hence, the possible moves – of adversaries. Most games in the argumentation literature are games of perfect information, while many assume complete information of the adversary, or don't care. For dialogues that are collaborative, seeking to find a joint truth[18], privacy/incomplete information would seem not to matter; for those designed to provide an operational characterization (or proof theory) for specific semantics[19], again it would seem that privacy does not matter. Many works seeking to apply game-theoretic solution concepts, such as Nash equilibria, to argumentation games [120; 115; 97; 50] assume players have complete information about an adversary's possible moves, since that is an underlying assumption of Nash equilibria. On the other hand, many argumentation games in the literature are incomplete information games, for example [121; 107; 116; 125; 27; 69].

One way of analysing argumentation games of incomplete information is to frame them as *Bayesian extensive games with observable actions* [106, chap. 12]: this is possible because every player observes the move of the other player and uncertainty only derives from an initial move of Chance that distributes private information (rules or arguments) among the players. 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, for example, in [120; 69]. Despite this fact, however, complexity results for Bayesian games are far from encouraging (see [61] for games of strategy). Indeed, it seems that considerations similar to those presented by [34] can be applied to argument games: the calculation of the perfect Bayesian equilibrium solution can

[18] Such dialogues are known as *inquiry* dialogues [138].

[19] Examples of such work are [136; 3; 100].

be tremendously complex due to both the size of the strategy space (as a function of the size of the game tree, and it can be computationally hard to compute it [40]), and the dependence between variables representing strategies and players' beliefs.

Many works, for example [119; 70] (and see [127; 24] for more discussion), have addressed the development of a model of the adversary, which can help in developing heuristics for choosing a particular move. Such work does not change the worst-case complexity of making a move, which is NP-hard or worse (see Table 2). Furthermore, even with full knowledge of the adversary, the problem of developing a strategy to beat the adversary is PSPACE-complete (Theorem 7.2).

As mentioned earlier, some work [116; 122] considers hiding arguments (that is, playing an argument a_1 that you know is defeated by a_2, but keeping a_2 private) to be dishonest or even lying. However, in a game of incomplete knowledge a player does not know which arguments hold in the omniscient argumentation framework, so this attitude seems harsh. In any case, our focus is on strategic arguing, where hiding arguments is acceptable. Those works also address "bullshitting" [56] (the introduction of arguments that the player does not know), which is not acceptable in strategic argumentation. We assign to the adjudicator the responsibility for rejecting such arguments. [116] shows that, for their single simultaneous move game, honesty is the best policy only in very restrictive circumstances. [122] identifies some cases in which a player can detect a dishonest adversary, while [107] show that, as the players play more games the probability of a lie being caught by the adversary approaches 1. Apart from these works, which might be considered as addressing corruption isolated to a single player, there seems no discussion of corruption in formal argumentation prior to [88]. [126] address "argumentational integrity", but this refers to fairness in the performance of general argumentation; they do, however, agree that "pretence of truth" is unfair, and would also consider hidden arguments as "insincere contributions".

A majority of the (persuasion) dialogue and argumentation literature takes the perspective of Dung, which sees arguments as monadic elements. There, arguments are typically abstract: the players know such arguments, can propose one (or a set) of them during a turn of the game, but the players do not know their internal structure. Although for many applications this perspective is admissible and gives good benefits in simplifying the problem, in some cases it results in an oversimplification. Anyway, restricting to abstract argumentation does not reduce the complexity of the problems, in general. We have seen in Section 7.5 that hardness results at the abstract level can be extended to the concrete level. Thus, it seems that the complexity of the problems largely comes from the problems themselves (including semantics and strategic aims) and not from the level of detail of the arguments.

Strategic argumentation can be considered a specific form of collective argumen-

tation [25] (and judgement aggregation), where different argumentation frameworks contribute to a combined judgement on the arguments. This topic is usually considered in the context of collaboration, but some work considers self-interested agents [27; 77]. Strategic argumentation is clearly a framework-wise approach, in the classification of [25], where argument frameworks are combined, and arguments then evaluated in the result. See Chapter 4 [18] of this handbook for additional discussion of this topic from a computational social choice perspective.

An approach to argumentation of interest for strategic argumentation is *probabilistic argumentation*. We refer the readers to Chapter 7 of this volume [73] for an in-depth discussion of this topic. Under the constellations approach to probabilistic argumentation, the key idea is that the existence (or, perhaps, validity) of arguments and attacks is unknown, but there is a probability distribution function describing the likelihood of different possibilities. Such an approach could be a useful refinement for strategic argumentation, allowing the replacement of a complete unknown (the adversary's arguments) with a more detailed model of the adversary. This might provide the basis for a player to choose among different moves.

Within the framework proposed in [80], probabilities are used to represent the likelihood that arguments and attacks exist. This defines a probability distribution over all possible worlds, where each possible world is an abstract argumentation framework consisting of some subset of the arguments and attacks. Extensions arise, as usual, for a possible world, by applying any of various semantics. In [80; 52], the authors tackle the probabilistic counterpart of the problem $\text{VER}^\sigma(S)$, that is, the problem $\text{PROB}^\sigma_\mathcal{F}(S)$ of computing the probability $Prs^\sigma_\mathcal{F}(S)$ that a set S of arguments is an extension according to a given semantics σ, given a probabilistic argumentation framework $\mathcal{F}$. [80] suggested that computing the exact value of probability $Prs^\sigma_\mathcal{F}(S)$ requires exponential time, and employed a Monte-Carlo simulation approach to approximate $\text{PROB}^\sigma_\mathcal{F}(S)$. However, as far as the admissible and stable semantics are concerned, [52]'s results show that the exact value of $Prs^\sigma_\mathcal{F}(S)$ can be determined in polynomial time, without enumerating the possible worlds. Nevertheless, in general the number of extensions is potentially exponential and, for other semantics, the problem is intractable. Consequently, it seems likely that many of the problems arising in strategic probabilistic argumentation will also be difficult.

Finally, there are some works that might appear to be addressing strategic argumentation, but have only weak relevance to the topic. *Strategic manoeuvring* was introduced in [133] to bridge the gap between dialectical and rhetorical approaches to the study of argumentation [134]. It refers to "the efforts arguers make in argumentative discourse to reconcile aiming for rhetorical effectiveness with maintaining dialectical standards of reasonableness" [134]. It was introduced in the context of the pragma-dialectical theory of argumentation [132;

131], which focuses on analysis and evaluation of lingual argumentation. This theory is a much broader view of argumentation than we address here. Nevertheless, there might be links between strategic manoeuvring and strategic argumentation applied to value-based or audience-based argumentation frameworks [19].

We have already mentioned [126], which addresses ethics of lingual argumentative communication. It proposes standards for lingual argumentation, under the title *argumentational integrity*, and develops a taxonomy of these standards. The standards address rhetoric rather than the relation between arguments, and the notion of integrity does not include corruption (except to the extent already discussed in Section 7.1).

Despite the title, [46] analyzes a very different scenario than we do here. In that work, a decision-maker consults an expert, who possibly has an ulterior motive, about deciding between two alternatives. For example, a customer consulting a camera salesman about which camera to buy. The expert has all the arguments (which are informal) for both alternatives, and the decision-maker has none. The game is modelled probabilistically, and the paper performs an equilibrium analysis. Apart from the words "strategic argumentation" and the possibility of a self-interested player, there is no relationship between this work and the work on strategic argumentation presented here.

9 Future Directions

There are multiple avenues for further research in this area.

- The NP-completeness results in Section 5.1 apply to a wide variety of logics whose inference problem can be solved in polynomial time. Other logics, such as those in [21], that have a harder inference problem might result in complexities higher in the polynomial hierarchy. An analysis of such cases could extend the existing results.

- Structured argumentation theories can generate a large number of arguments, possibly infinitely many. This prevents applying the results of Section 7 to structured argumentation directly. For example, we used a different method to prove Theorem 5.8. What is needed is to find a polynomially-sized argumentation framework that is equivalent to the generated argumentation framework for the semantics of interest.

- In this chapter we have focused on a competitive situation, where the two players' aims are inconsistent. However, the basics of strategic argumentation also apply

when the player's aims are consistent. In this case, strategic argumentation represents a crude adversarial negotiation. It is worth exploring how concepts from strategic argumentation can be used to analyse such negotiations, both in strategic argumentation games and in other negotiation games.

- Work has focused on two-player games of strategic argumentation. However, there are often more than two stakeholders in an adjudication, and so it would be interesting to see how strategic argumentation can be extended to more players. Among the many issues that would need to be addressed are: the protocol for turn-taking, the criterion for terminating the game, and the possibility of some players cooperating to construct an argument that none of them could construct individually. There is discussion of multi-party dialogues in [37; 99; 127]. In general, game play would appear to be more complex because of the potential for shifting alliances between players, and because players might not be compelled to make a move at each opportunity. Corruption might also be more complicated.

- In current work, the players' aims are implicitly assumed to be known and fixed. In some scenarios this might be realistic. However, there are scenarios where the motivations of a player are unclear, and/or may change over the course of argumentation. For example, a defence lawyer might begin with a "not guilty" aim but, if the trial is going badly, change tack to instead aim at a mis-trial. Thus, the extension of strategic argumentation to consider aims as possibly private and flexible/changeable is an interesting one.

- In the treatment of strategic abstract argumentation, the most prominent semantics for Dung's framework have been addressed, but there remain many semantics in the literature for which resistance to corruption is unknown. In addition, the treatment of the subset-minimality standard remains to be done for most semantics.

- The treatment of espionage assumes that full knowledge of an adversary's arguments is obtained. Perhaps the illicit gain of only some knowledge is more realistic. How can this framework be extended to cases where only partial knowledge is obtained? The work of [39] could be a first step in this direction. That paper represents partial knowledge and determines whether a player has the ability to force a desired outcome. However, it will need much expansion, as it only addresses Existential and Universal outcomes, and only for the stable semantics; assumes that the player's control arguments cannot be attacked by partially-known attacks; and does not consider multiple moves.

- Although standards are insufficient to make corruption visible, they can also be useful in guiding heuristic approaches to playing strategic argumentation games. For example, the subset-minimality standard prevents a player needlessly creating an opening for the adversary. ([105] employ this as a heuristic in a different dialogue game than the one we have presented.) Thus, it would be helpful to identify more standards, especially those that can be incorporated in heuristics or used to improve a heuristic move.

- The brief discussion of argument retraction in Section 8 deserves expansion. Strategic argumentation with retraction would seem to produce an outcome that is less arbitrary than without retraction, but perhaps the strategic element would be much diminished. Argument retraction would need to be restricted in some way, or an explicit termination rule introduced, otherwise a losing player might be able to prevent termination by repeatedly retracting arguments and then replaying them. Treating such retraction as a disavowal of some or all of the backtracked arguments (i.e. a commitment not to use those arguments in the remainder of the game) might temper the power of retraction and lead to a richer game.

- The notion of resistance to corruption we discussed is based on worst-case complexity, but this is sometimes not reflective of the difficulty of problems that arise in practice. An empirical comparison of the difficulty of solving the problems in practice and a study of approximation algorithms for these problems are needed.

- As observed in subsection 6.1, it can be worthwhile to consider an adjudicator as part of a strategic argumentation game. In this case we might consider whether the adjudicator can be subject to corruption. If the role of the adjudicator is simply to enforce the consequences arrived at by the players then there is nothing *in the game* that allows us to detect corruption.

  However, if we assume that the adjudicator chooses the semantics under which the game will be adjudicated, we have an action by the adjudicator that can be subject to analysis. This leads to quite different games, especially if the adjudicator changes the semantics *during* the playing of the game. While this appears to be rather Kafkaesque, it might be somewhat reflective of some situations where the judiciary can be influenced by other arms of government. The adjudicator then has both the choice of semantics to impose, and the choice of timing of this move. More realistically, [110] presents a game where the adjudicator plays an active role, based on a detailed model of legal procedure.

Perhaps that model is a base on which corruption of adjudication can be investigated.

10 Conclusions

Strategic argumentation is a primarily adversarial approach to dialogue games with incomplete information. It reflects aspects of legal argument. The idea can be applied at a concrete level, as we have demonstrated using defeasible logic rules as the basis for arguments, and at an abstract level, which was demonstrated using Dung's argumentation system.

The key element of strategic argumentation games is each player re-establishing their aim at the end of their turn. The details of the argument framework are not needed at this level of abstraction, only that they can be used to define a notion of acceptance/aim achievement. Consequently, we have a formulation of strategic argumentation that applies to Dung's notion of argumentation framework [41], but also to bipolar argumentation frameworks [33], abstract argumentation frameworks with sets of attacking arguments [101; 55][20], and preference-based argumentation frameworks [3; 17]. If, in the dialogue game $(\mathcal{A}, \mathcal{R})$, we extend $\mathcal{R}$ beyond simply relations on $\mathcal{A}$ then we can have strategic argumentation on constrained argumentation frameworks [35], weighted argument systems [42], abstract dialectical frameworks [30], and probabilistic argumentation frameworks [80], and the ideas might well be applicable to other forms of argumentation framework. Similarly, the ideas of strategic argumentation apply to semantics other than Dung-style semantics.

We have also demonstrated how the strategic argumentation framework can be used to address issues of corruption, even when the corrupt behaviour is motivated by rewards extrinsic to the game. We have not much addressed the strategies that a player might employ when playing a strategic argumentation game, although the study of standards in Section 7 provides some guidelines. More information on that topic can be found in Section 5.2 of Chapter 9 [24] in this handbook.

Acknowledgments

We thank the reviewers for their comments, which helped to improve this chapter. Michael Maher has an adjunct position at Griffith University and an honorary position at UNSW.

[20] We have already addressed argumentation with sets of attacking arguments in a non-abstract setting, in Section 5.

References

[1] Leila Amgoud. A replication study of semantics in argumentation. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019*, pages 6260–6266. ijcai.org, 2019.

[2] Leila Amgoud, Lianne Bodenstaff, Martin Caminada, Peter McBurney, Simon Parsons, Henry Prakken, Jelle van Veenen, and Gerard Vreeswijk. Final review and report on formal argumentation system. Technical report, 2006.

[3] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Ann. Math. Artif. Intell.*, 34(1-3):197–215, 2002.

[4] Grigoris Antoniou. A Discussion of Some Intuitions of Defeasible Reasoning. In George Vouros and Themistoklis Panayiotopoulos, editors, *Methods and Applications of Artificial Intelligence*, volume 3025 of *Lecture Notes in Computer Science*, pages 311–320. Springer Berlin / Heidelberg, 2004.

[5] Grigoris Antoniou, David Billington, Guido Governatori, and Michael J. Maher. A flexible framework for defeasible logics. In *AAAI/IAAI*, pages 405–410, 2000.

[6] Grigoris Antoniou, David Billington, Guido Governatori, and Michael J. Maher. Representation results for defeasible logic. *ACM Trans. Comput. Log.*, 2(2):255–287, 2001.

[7] Grigoris Antoniou, David Billington, Guido Governatori, and Michael J. Maher. Representation results for defeasible logic. *ACM Transactions on Computational Logic*, 2(2):255–286, 2001.

[8] Grigoris Antoniou, David Billington, Guido Governatori, Michael J. Maher, and Andrew Rock. A family of defeasible reasoning logics and its implementation. In *Proc. ECAI-2000*, pages 459–463, 2000.

[9] Grigoris Antoniou, David Billington, and Michael J. Maher. On the analysis of regulations using defeasible rules. In *32nd Annual Hawaii International Conference on System Sciences (HICSS-32)*, 1999.

[10] Grigoris Antoniou, Thomas Skylogiannis, Antonis Bikakis, Martin Doerr, and Nick Bassiliades. Dr-brokering: A semantic brokering system. *Knowledge-Based Systems*, 20(1):61–72, 2007.

[11] Joseph M. Barbato. Scotland's bastard verdict: Intermediacy and the unique three-verdict system. *Indiana International & Comparative Law Review*, 15:543–582, 2005.

[12] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. An introduction to argumentation semantics. *Knowledge Eng. Review*, 26(4):365–410, 2011.

[13] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. Abstract argumentation frameworks and their semantics. In Pietro Baroni; Dov Gabbay; Massimiliano Giacomin; Leendert van der Torre, editor, *Handbook on Formal Argumentation*, volume 1, pages 688–767. College Publications, February 2018.

[14] Pietro Baroni, Paul E. Dunne, and Massimiliano Giacomin. On extension counting problems in argumentation frameworks. In Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Guillermo Ricardo Simari, editors, *Computational Models*

of Argument: Proceedings of COMMA 2010, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 63–74. IOS Press, 2010.

[15] Pietro Baroni, Antonio Rago, and Francesca Toni. How many properties do we need for gradual argumentation? In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 1736–1743, 2018.

[16] John J. Bartholdi, Craig A. Tovey, and Michael A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989.

[17] Dorothea Baumeister, Daniel Neugebauer, and Jörg Rothe. Verification in attack-incomplete argumentation frameworks. In Toby Walsh, editor, *Algorithmic Decision Theory*, pages 341–358. Springer International Publishing, 2015.

[18] Dorothea Baumeister, Daniel Neugebauer, and Jörg Rothe. Collective acceptability in abstract argumentation. In Dov Gabbay, Massimiliano Giacomin, Guillermo R. Simari, and Matthias Thimm, editors, *Handbook of Formal Argumentation*, volume 2, chapter 4. College Publications, 2021.

[19] Trevor J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *J. Log. Comput.*, 13(3):429–448, 2003.

[20] Salem Benferhat, Didier Dubois, and Henri Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In David Heckerman and E. H. Mamdani, editors, *UAI '93: Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence*, pages 411–419. Morgan Kaufmann, 1993.

[21] David Billington. A defeasible logic for clauses. In *AI 2011: Advances in Artificial Intelligence - 24th Australasian Joint Conference, Proceedings*, pages 472–480, 2011.

[22] David Billington. *Factual and Plausible Reasoning*, volume 81 of *Studies in Logic*. College Publications, 2019.

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

[24] Elizabeth Black, Nicolas Maudet, and Simon Parsons. Argumentation-based dialogue. In Dov Gabbay, Massimiliano Giacomin, Guillermo R. Simari, and Matthias Thimm, editors, *Handbook of Formal Argumentation*, volume 2, chapter 9. College Publications, 2021.

[25] Gustavo Adrian Bodanza, Fernando Tohmé, and Marcelo Auday. Collective argumentation: A survey of aggregation issues around argumentation frameworks. *Argument & Computation*, 8(1):1–34, 2017.

[26] Andrei Bondarenko, Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artif. Intell.*, 93:63–101, 1997.

[27] Elise Bonzon and Nicolas Maudet. On the outcomes of multiparty persuasion. In *Argumentation in Multi-Agent Systems - 8th International Workshop, ArgMAS 2011*,

volume 7543 of *Lecture Notes in Computer Science*, pages 86–101. Springer, 2012.

[28] Annalisa Bossi, Michele Bugliesi, Maurizio Gabbrielli, Giorgio Levi, and Maria Chiara Meo. Differential logic programming. In *Conference Record of the Twentieth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, pages 359–370, 1993.

[29] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[30] Gerhard Brewka and Stefan Woltran. Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010, Toronto, Ontario, Canada, May 9-13, 2010*, 2010.

[31] Martin Caminada, Samy Sá, João Alcântara, and Wolfgang Dvořák. On the equivalence between logic programming semantics and argumentation semantics. *Int. J. Approx. Reasoning*, 58:87–111, 2015.

[32] Claudette Cayrol, Sylvie Doutre, Marie-Christine Lagasquie-Schiex, and Jérôme Mengin. "Minimal defence": a refinement of the preferred semantics for argumentation frameworks. In *9th International Workshop on Non-Monotonic Reasoning (NMR 2002), Proceedings*, pages 408–415, 2002.

[33] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 8th European Conference, ECSQARU 2005, Proceedings*, pages 378–389, 2005.

[34] Georgios Chalkiadakis and Craig Boutilier. Coalitional bargaining with agent type uncertainty. In Manuela M. Veloso, editor, *IJCAI*, pages 1227–1232, 2007.

[35] Sylvie Coste-Marquis, Caroline Devred, and Pierre Marquis. Constrained argumentation frameworks. In *Proceedings, Tenth International Conference on Principles of Knowledge Representation and Reasoning*, pages 112–122, 2006.

[36] Joseph Devereux and Chris Reed. Strategic argumentation in rigorous persuasion dialogue. In *Argumentation in Multi-Agent Systems, 6th International Workshop, ArgMAS 2009*, pages 94–113, 2009.

[37] Frank Dignum and Gerard Vreeswijk. Towards a testbed for multi-party dialogues. In Frank Dignum, editor, *Advances in Agent Communication, International Workshop on Agent Communication Languages, ACL 2003, Melbourne, Australia, July 14, 2003*, volume 2922 of *Lecture Notes in Computer Science*, pages 212–230. Springer, 2003.

[38] Yannis Dimopoulos and Antonis C. Kakas. Logic programming without negation as failure. In *Proceedings of the 1995 International Symposium on Logic Programming*, pages 369–384, 1995.

[39] Yannis Dimopoulos, Jean-Guy Mailly, and Pavlos Moraitis. Control argumentation frameworks. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)*, pages 4678–4685. AAAI Press, 2018.

[40] Yannis Dimopoulos, Bernhard Nebel, and Francesca Toni. On the computational

complexity of assumption-based argumentation for default reasoning. *Artif. Intell.*, 141(1/2):57–78, 2002.

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

[42] Paul E. Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael J. Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artif. Intell.*, 175(2):457–486, 2011.

[43] Paul E. Dunne and Michael Wooldridge. Complexity of abstract argumentation. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 85–104. Springer, 2009.

[44] Wolfgang Dvořák. On the complexity of computing the justification status of an argument. In Sanjay Modgil, Nir Oren, and Francesca Toni, editors, *Theorie and Applications of Formal Argumentation - First International Workshop, TAFA*, volume 7132 of *Lecture Notes in Computer Science*, pages 32–49. Springer, 2011.

[45] Wolfgang Dvořák and Paul E. Dunne. Computational problems in formal argumentation and their complexity. *IfCoLog Journal of Logics and their Applications*, 4(8), 2017.

[46] Wioletta Dziuda. Strategic argumentation. *J. Econ. Theory*, 146(4):1362–1397, 2011.

[47] Kathleen M. Eisenhardt. Agency theory: An assessment and review. *The Academy of Management Review*, 14(1):57–74, 1989.

[48] Jenny Eriksson Lundström, Guido Governatori, Subhasis Thakur, and Vineet Padmanabhan. An asymmetric protocol for argumentation games in defeasible logic. In Aditya Ghose and Guido Governatori, editors, *10 Pacific Rim International Workshop on Multi-Agents*, volume 5044 of *LNAI*, pages 219–231, Heidelberg, 2008. Springer.

[49] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Common Knowledge Revisited*. MIT Press, 2003.

[50] Xiuyi Fan and Francesca Toni. On the interplay between games, argumentation and dialogues. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems, Singapore, May 9-13, 2016*, pages 260–268, 2016.

[51] Arthur M. Farley and Kathleen Freeman. Burden of proof in legal argumentation. In *Proceedings of the Fifth International Conference on Artificial Intelligence and Law, ICAIL '95*, pages 156–164, 1995.

[52] Bettina Fazzinga, Sergio Flesca, and Francesco Parisi. On the complexity of probabilistic abstract argumentation. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, 2013.

[53] Johannes Klaus Fichte, Markus Hecher, and Arne Meier. Counting complexity for reasoning in abstract argumentation. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI*, pages 2827–2834. AAAI Press, 2019.

[54] Johannes C. Flieger. Relevance and minimality in systems of defeasible argumentation. Technical report, Imperial College, London, 2002.

[55] Giorgos Flouris and Antonis Bikakis. A comprehensive study of argumentation frame-

works with sets of attacking arguments. *Int. J. Approx. Reason.*, 109:55–86, 2019.

[56] Harry G. Frankfurt. *On Bullshit*. Princeton University Press, 2005.

[57] Dinesh Garg, Yadati Narahari, and Sujit Gujar. Foundations of mechanism design: A tutorial. *Sadhana*, 33(2), 2008.

[58] Thomas F. Gordon. The pleadings game: An exercise in computational dialectics. *Artif. Intell. Law*, 2:239–292, December 1994.

[59] Thomas F. Gordon, Henry Prakken, and Douglas Walton. The Carneades model of argument and burden of proof. *Artif. Intell.*, 171(10-15):875–896, 2007.

[60] Thomas F. Gordon and Douglas Walton. Proof burdens and standards. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 239–260. Springer, 2009.

[61] Georg Gottlob, Gianluigi Greco, and Toni Mancini. Complexity of pure equilibria in bayesian games. In Manuela M. Veloso, editor, *IJCAI*, pages 1294–1299, 2007.

[62] Guido Governatori. On the relationship between Carneades and Defeasible Logic. In *The 13th International Conference on Artificial Intelligence and Law, Proceedings of the Conference*, pages 31–40, 2011.

[63] Guido Governatori, Marlon Dumas, Arthur H.M. ter Hofstede, and Phillipa Oaks. A formal approach to protocols and strategies for (legal) negotiation. In Henry Prakken, editor, *Procedings of the 8th International Conference on Artificial Intelligence and Law*, pages 168–177. IAAIL, ACM Press, 2001.

[64] Guido Governatori and Michael J. Maher. Annotated defeasible logic. *Theory Pract. Log. Program.*, 17(5-6):819–836, 2017.

[65] Guido Governatori, Michael J. Maher, Grigoris Antoniou, and David Billington. Argumentation semantics for defeasible logics. *Journal of Logic and Computation*, 14(5):675–702, 2004.

[66] Guido Governatori, Francesco Olivieri, Antonino Rotolo, and Simone Scannapieco. Computing strong and weak permissions in defeasible logic. *Journal of Philosophical Logic*, 42(6):799–829, 2013.

[67] Guido Governatori, Francesco Olivieri, Simone Scannapieco, Antonino Rotolo, and Matteo Cristani. Strategic argumentation is NP-complete. In *Proc. ECAI 2014*. IOS Press, 2014.

[68] Guido Governatori and Antonino Rotolo. Changing legal systems: legal abrogations and annulments in defeasible logic. *Logic Journal of IGPL*, 18(1):157–194, 2010.

[69] Davide Grossi and Wiebe van der Hoek. Audience-based uncertainty in abstract argument games. In *IJCAI'13*, pages 143–149. AAAI Press, 2013.

[70] Christos Hadjinikolis, Yiannis Siantos, Sanjay Modgil, Elizabeth Black, and Peter McBurney. Opponent modelling in persuasion dialogues. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 2013.

[71] Emmanuel Hadoux, Aurélie Beynier, Nicolas Maudet, Paul Weng, and Anthony Hunter. Optimization of probabilistic argumentation with markov decision models. In Qiang Yang and Michael J. Wooldridge, editors, *Proceedings of the Twenty-Fourth Interna-*

tional Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, pages 2004–2010. AAAI Press, 2015.

[72] Mustafa Hashmi, Guido Governatori, Ho-Pun Lam, and Moe Thandar Wynn. Are we done with business process compliance: state of the art and challenges ahead. *Knowl. Inf. Syst.*, 57(1):79–133, 2018.

[73] Anthony Hunter, Sylwia Polberg, Nico Potyka, Tjitze Rienstra, and Matthias Thimm. Probabilistic argumentation: A survey. In Dov Gabbay, Massimiliano Giacomin, Guillermo R. Simari, and Matthias Thimm, editors, *Handbook of Formal Argumentation*, volume 2, chapter 7. College Publications, 2021.

[74] Mohammad Badiul Islam and Guido Governatori. RuleRS: a rule-based architecture for decision support systems. *Artif. Intell. Law*, 26(4):315–344, 2018.

[75] David S. Johnson. A catalog of complexity classes. In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity*, pages 67–161. Elsevier, 1990.

[76] Antonis C. Kakas, Pavlos Moraitis, and Nikolaos I. Spanoudakis. *GORGIAS*: Applying argumentation. *Argument & Computation*, 10(1):55–81, 2019.

[77] Dionysios Kontarinis, Elise Bonzon, Nicolas Maudet, and Pavlos Moraitis. On the use of target sets for move selection in multi-agent debates. In *ECAI 2014 - 21st European Conference on Artificial Intelligence*, pages 1047–1048, 2014.

[78] Els Laenens and Dirk Vermeir. A fixpoint semantics for ordered logic. *J. Log. Comput.*, 1(2):159–185, 1990.

[79] Ho-Pun Lam, Guido Governatori, and Régis Riveret. On aspic$^+$ and defeasible logic. In *Computational Models of Argument - Proceedings of COMMA 2016*, pages 359–370, 2016.

[80] Hengfei Li, Nir Oren, and Timothy J. Norman. Probabilistic argumentation frameworks. In Sanjay Modgil, Nir Oren, and Francesca Toni, editors, *Theorie and Applications of Formal Argumentation - First International Workshop, TAFA 2011. Revised Selected Papers*, volume 7132 of *Lecture Notes in Computer Science*, pages 1–16. Springer, 2012.

[81] Paul Lorenzen and Kuno Lorenz. *Dialogische Logik*. Darmstadt, 1978.

[82] Ronald Prescott Loui. Process and policy: Resource-bounded nondemonstrative reasoning. *Computational Intelligence*, 14(1):1–38, 1998.

[83] Michael J. Maher. Propositional defeasible logic has linear complexity. *Theory and Practice of Logic Programming*, 1(6):691–711, 2001.

[84] Michael J. Maher. Relative expressiveness of defeasible logics. *Theory and Practice of Logic Programming*, 12(4-5):793–810, 2012.

[85] Michael J. Maher. Relative expressiveness of defeasible logics II. *Theory and Practice of Logic Programming*, 13:579–592, 2013.

[86] Michael J. Maher. Relative Expressiveness of Well-Founded Defeasible Logics. In Stephen Cranefield and Abhaya Nayak, editors, *AI 2013: Advances in Artificial Intelligence*, volume 8272 of *Lecture Notes in Computer Science*, pages 338–349. Springer International Publishing, 2013.

[87] Michael J. Maher. Comparing defeasible logics. In Torsten Schaub, Gerhard Friedrich,

and Barry O'Sullivan, editors, *ECAI 2014 - 21st European Conference on Artificial Intelligence*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 585–590. IOS Press, 2014.

[88] Michael J. Maher. Complexity of exploiting privacy violations in strategic argumentation. In Duc Nghia Pham and Seong-Bae Park, editors, *PRICAI 2014: Trends in Artificial Intelligence - 13th Pacific Rim International Conference on Artificial Intelligence*, volume 8862 of *Lecture Notes in Computer Science*, pages 523–535. Springer, 2014.

[89] Michael J. Maher. Corrupt strategic argumentation: The ideal and the naive. In *AI 2016: Advances in Artificial Intelligence*, pages 17–28, 2016.

[90] Michael J. Maher. Resistance to corruption of general strategic argumentation. In *Proc. Int. Conf. Principles and Practice of Multi-Agent Systems*, pages 61–75, 2016.

[91] Michael J. Maher. Resistance to corruption of strategic argumentation. In *AAAI Conference on Artificial Intelligence*, pages 1030–1036, 2016.

[92] Michael J. Maher. Corruption and audit in strategic argumentation. Technical report, Reasoning Research Institute, 2017.

[93] Michael J. Maher. Relating concrete defeasible reasoning formalisms and abstract argumentation. *Fundam. Inform.*, 155(3):233–260, 2017.

[94] Michael J. Maher, Ilias Tachmazidis, Grigoris Antoniou, Stephen Wade, and Long Cheng. Rethinking defeasible reasoning: A scalable approach. *Theory and Practice of Logic Programming*, 20(4):552–586, 2020.

[95] Frederick Maier. Interdefinability of defeasible logic and logic programming under the well-founded semantics. *Theory and Practice of Logic Programming*, 13:107–142, 2013.

[96] Frederick Maier and Donald Nute. Well-founded semantics for defeasible logic. *Synthese*, 176(2):243–274, 2010.

[97] P. Matt and F. Toni. A game-theoretic measure of argument strength for abstract argumentation. In *JELIA 2008*, volume 5293 of *LNCS*, pages 285–297. Springer, 2008.

[98] Mohamed Mbarki, Jamal Bentahar, and Bernard Moulin. Specification and complexity of strategic-based reasoning using argumentation. In *Argumentation in Multi-Agent Systems, Third International Workshop, ArgMAS 2006, Hakodate, Japan, May 8, 2006, Revised Selected and Invited Papers*, pages 142–160, 2006.

[99] Peter McBurney and Simon Parsons. Dialogue games for agent argumentation. In Guillermo Ricardo Simari and Iyad Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 261–280. Springer, 2009.

[100] Sanjay Modgil and Martin Caminada. Proof theories and algorithms for abstract argumentation frameworks. In *Argumentation in Artificial Intelligence*, pages 105–129. Springer, 2009.

[101] Søren Holbech Nielsen and Simon Parsons. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Argumentation in Multi-Agent Systems, Third International Workshop, ArgMAS 2006*, pages 54–73, 2006.

[102] Donald Nute. Defeasible reasoning: A philosophical analysis in Prolog. In James H. Fetzer, editor, *Aspects of Artificial Intelligence*, volume 1 of *Studies in Cognitive Systems*. Springer, 1988.

[103] Donald Nute. Defeasible logic. In Dov M. Gabbay, Chris J. Hogger, and J. Allen Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3. Oxford University Press, 1994.

[104] Donald Nute. Defeasible logic: Theory, Implementation and Applications. In *Proceedings of the 14th International Conference on Applications of Prolog (INAP 2001)*, pages 151–169. Springer, Berlin, 2001.

[105] Nir Oren, Timothy J. Norman, and Alun D. Preece. Loose lips sink ships: a heuristic for argumentation. In *Proceedings of the Third International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2006)*, pages 121–134, 2006.

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

[107] Simon Parsons and Elizabeth Sklar. How agents alter their beliefs after an argumentation-based dialogue. In *Argumentation in Multi-Agent Systems, Second International Workshop, ArgMAS 2005*, pages 297–312, 2005.

[108] Henry Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:2001, 2000.

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

[110] Henry Prakken. A formal model of adjudication dialogues. *Artif. Intell. Law*, 16(3):305–328, 2008.

[111] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124, 2010.

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

[113] Henry Prakken and Giovanni Sartor. Modelling reasoning with precedents in a formal dialogue game. *Artif. Intell. Law*, 6(2-4):231–287, 1998.

[114] Henry Prakken and Giovanni Sartor. Law and logic: A review from an argumentation perspective. *Artif. Intell.*, 227:214–245, 2015.

[115] Ariel D. Procaccia and Jeffrey S. Rosenschein. Extensive-form argumentation games. In Marie-Pierre Gleizes, Gal A. Kaminka, Ann Nowé, Sascha Ossowski, Karl Tuyls, and Katja Verbeeck, editors, *EUMAS 2005 - Proceedings of the Third European Workshop on Multi-Agent Systems*, pages 312–322. Koninklijke Vlaamse Academie van Belie voor Wetenschappen en Kunsten, 2005.

[116] Iyad Rahwan, Kate Larson, and Fernando A. Tohmé. A characterisation of strategy-proofness for grounded argumentation semantics. In *IJCAI*, pages 251–256, 2009.

[117] Iyad Rahwan, Sarvapali D. Ramchurn, Nicholas R. Jennings, Peter Mcburney, Simon Parsons, and Liz Sonenberg. Argumentation-based negotiation. *The Knowledge Engineering Review*, 18(4):343–375, 2003.

[118] Daniel M. Reeves, Michael P. Wellman, and Benjamin N. Grosof. Automated negotiation

from declarative contract descriptions. *Computational Intelligence*, 18(4):482–500, 2002.

[119] Tjitze Rienstra, Matthias Thimm, and Nir Oren. Opponent models with uncertainty for strategic argumentation. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 2013.

[120] Régis Riveret, Henry Prakken, Antonino Rotolo, and Giovanni Sartor. Heuristics in argumentation: A game theory investigation. In *COMMA 2008*, pages 324–335. IOS Press, 2008.

[121] Bram Roth, Régis Riveret, Antonino Rotolo, and Guido Governatori. Strategic argumentation: a game theoretical investigation. In *The Eleventh International Conference on Artificial Intelligence and Law, Proceedings of the Conference*, pages 81–90, 2007.

[122] Chiaki Sakama. Dishonest arguments in debate games. In *Computational Models of Argument - Proceedings of COMMA 2012, Vienna, Austria, September 10-12, 2012*, pages 177–184, 2012.

[123] Chiaki Sakama. Debate games in logic programming. In *Declarative Programming and Knowledge Management - Declarative Programming Days, KDPD 2013*, pages 185–201, 2013.

[124] K. Satoh and K. Takahashi. A semantics of argumentation under incomplete information. In *Proceedings of Jurisn 2011*, 2011.

[125] Ken Satoh and Kazuko Takahashi. A semantics of argumentation under incomplete information. In *JURISIN*, pages 86–97, 2011.

[126] Margrit Schreier, Norbert Groeben, and Ursula Christmann. "That's Not Fair!" argumentational integrity as an ethics of argumentative communication. *Argumentation*, 9:267–289, 1995.

[127] Matthias Thimm. Strategic argumentation in multi-agent systems. *Künstliche Intell.*, 28(3):159–168, 2014.

[128] Matthias Thimm and Alejandro Javier García. On strategic argument selection in structured argumentation systems. In *Argumentation in Multi-Agent Systems - 7th International Workshop, ArgMAS 2010*, pages 286–305, 2010.

[129] David S. Touretzky. *The Mathematics of Inheritance Systems*. Morgan Kaufmann, 1986.

[130] David S. Touretzky, John F. Horty, and Richmond H. Thomason. A Clash of Intuitions: The Current State of Nonmonotonic Multiple Inheritance Systems. In *Proceedings of the 10th international joint conference on Artificial intelligence (IJCAI'87)*, pages 476–482, San Francisco, CA, USA, 1987. Morgan Kaufmann Publishers Inc.

[131] Frans H. van Eemeren, editor. *Reasonableness and Effectiveness in Argumentative Discourse, Fifty Contributions to the Development of Pragma-Dialectics*, volume 27 of *Argumentation Library*. Springer, 2015.

[132] Frans H. van Eemeren and R. Grootendorst, editors. *A Systematic Theory of Argumentation. The Pragma-Dialectical Approach*. Cambridge University Press, 2004.

[133] Frans H. van Eemeren and Peter Houtlosser. Strategic maneuvering: Maintaining a

delicate balance. In Frans H. van Eemeren and Peter Houtlosser, editors, *Dialectic and Rhetoric: The Warp and Woof of Argumentation Analysis*. Kluwer Academic Publishers, 2002.

[134] Frans H. van Eemeren and Peter Houtlosser. Strategic maneuvering: A synthetic recapitulation. *Argumentation*, 20:381–392, 2006.

[135] Bart Verheij. DefLog: on the logical interpretation of prima facie justified assumptions. *J. Log. Comput.*, 13(3):319–346, 2003.

[136] Gerard Vreeswijk and Henry Prakken. Credulous and sceptical argument games for preferred semantics. In *Logics in Artificial Intelligence, European Workshop, JELIA 2000*, volume 1919 of *Lecture Notes in Computer Science*, pages 239–253. Springer, 2000.

[137] Klaus W. Wagner. The complexity of combinatorial problems with succinct input representation. *Acta Inf.*, 23(3):325–356, 1986.

[138] Douglas Walton and Erik C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, 1995.

[139] Hui Wan, Benjamin N. Grosof, Michael Kifer, Paul Fodor, and Senlin Liang. Logic programming with defaults and argumentation theories. In *ICLP*, pages 432–448, 2009.

[140] Hui Wan, Michael Kifer, and Benjamin N. Grosof. Defeasibility in answer set programs with defaults and argumentation rules. *Semantic Web*, 6(1):81–98, 2015.

[141] Stefan Woltran. Abstract argumentation – all problems solved? presentation at ECAI-14, 2014. `https://www.dbai.tuwien.ac.at/staff/woltran/ecai2014.pdf`.

[142] Yining Wu and Martin Caminada. A labelling-based justification status of arguments. *Studies in Logic*, 3(4):12–29, 2010.

[143] Yining Wu and Mikolaj Podlaszewski. Implementing crash-resistance and non-interference in logic-based argumentation. *J. Log. Comput.*, 25(2):303–333, 2015.

 Received May 2021