DRAFT VERSION

Künstliche Intelligenz manuscript No.
(will be inserted by the editor)

Strategic Argumentation in Multi-Agent Systems

Matthias Thimm

Received: date / Accepted: date

Abstract Argumentation-based negotiation describes the process of decision-making in multi-agent systems through the exchange of arguments. If agents only have partial knowledge about the subject of a dialogue strategic argumentation can be used to exploit weaknesses in the argumentation of other agents and thus to persuade other agents of a specific opinion and reach a certain outcome. This paper gives an overview of the field of strategic argumentation and surveys recent works and developments. We provide a general discussion of the problem of strategic argumentation in multi-agent settings and discuss approaches to strategic argumentation, in particular strategies based on opponent models.

Keywords Argumentation · Multi-Agent System · Strategies

1 Introduction

Computational models of argumentation [6] are an intuitive means for formalizing commonsense reasoning. The basic building blocks for argumentation systems are arguments, i. e., pieces of information that derive a claim, and an attack relation, i. e., a directed relation that represents conflict between arguments. Several research fields on computational models of argumentation have emerged in recent years such as abstract argumentation [11], structured argumentation [14, 7, 27], semantical issues [3], and, in particular, dynamic aspects of argumentation [13] and argumentation in multi-agent systems [20]. Within the latter field of research, issues related to *strategic aspects of argumentation* have gained some interest and constitute an active sub-field. Strategic argumentation takes place in multi-agent systems where agents aim to reach a common understanding for decision-making or try to persuade other agents of some opinion. Consider the following example with two agents Anna and Bob discussing whether or not the moon-landing happened in 1969:

Anna: The pictures supposedly taken during the moon-landing cannot be authentic as several shadows are inconsistent. So the moon-landing did not happen in 1969.

Bob: Due to reflected light from the Earth, shadows may appear inconsistent but they are not.

Anna: But the American flag that was hissed by the astronauts, fluttered despite the lack of wind.

Bob: The flag did not flutter. Ripples on the flag originating from folding it made it seem to flutter on a picture.

The above dialogue exemplifies how an exchange of arguments can be used to reach a common consensus. These kinds of dialogues offer opportunities for strate-

Matthias Thimm
Institute for Web Science and Technologies
Universität Koblenz-Landau
Universitätsstr. 1, 56070 Koblenz, Germany
Tel.: +49-261-287-2715
Fax: +49-261-287-100-2715
E-mail: thimm@uni-koblenz.de

gic exploitation, in particular, when agents have knowledge about their opponents' skills and beliefs. For example, assume that Anna knows that Bob is not an expert on astronomical phenomena. Then she could bring forward the following argument:

Anna: The amount of Van Allen radiation the astronauts were exposed to during the trip would have been lethal.

In real-world settings for argumentation, there is usually no time to process all arguments to reach a consensus. In such a setting it would have a strategic advantage for Anna to put forward the above argument first, instead of the other ones. Then Bob may be convinced that Anna is right in claiming that the moon-landing did not happen.

This overview paper surveys recent developments in strategic argumentation. In particular, we discuss the problem of *mechanism design* [29–31,37,12], i. e., the problem of coming up with argumentation protocols and negotiation settings where strategic argumentation has no benefit and the best strategy for every agent is to truthfully report all their arguments. Most of the work on mechanism design up to now focuses on abstract argumentation [11] and there have been many technical results on characterizing certain *strategy-proof* settings for argumentation. However, most of these results come with quite strict assumptions such as perfect knowledge, conflict-free preferences of agents, and certain requirements on the topology of the arguments and their relations. Therefore, we also discuss concrete strategies for argumentation [37,35] and focus on strategies exploiting an *opponent model* [23,16,33,17]. An opponent model is a component in the belief state of an agent that reflects what this agent believes what another agent believes. It can be used in adversarial games to predict how an opponent would react when performing a certain action, i. e., putting forward some argument. By using such a model, imperfect knowledge of an opponent can be exploited by putting forward those arguments where the opponent is unlikely to win the dialogue.

The remainder of this overview paper is organized as follows. In Section 2 we present some foundations of computational models in argumentation, in particular on abstract argumentation. In Section 3 we provide a general overview on multi-agent settings of argumentation and we provide a simple formalization of argumentation games in Section 4. In Section 5 we discuss the issue of strategic argumentation, with a particular focus on strategic argumentation with opponent models in Section 6. In Section 7 we discuss further works on strategic argumentation and we conclude with a discussion in Section 8.

2 Models of Argumentation

Abstract argumentation frameworks [11] take a very simple view on argumentation as they do not presuppose any internal structure of an argument. Abstract argumentation frameworks only consider the interactions of arguments by means of an attack relation between arguments.

Definition 1 (Abstract Argumentation Framework) An *abstract argumentation framework* AF is a tuple $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ where Arg is a set of arguments and $\rightarrow$ is a relation $\rightarrow \subseteq \mathsf{Arg} \times \mathsf{Arg}$.

For reasons of simplicity we only consider finitary argumentation frameworks here, i. e., argumentation frameworks with a finite number of arguments. For two arguments $\mathcal{A}, \mathcal{B} \in \mathsf{Arg}$ the relation $\mathcal{A} \rightarrow \mathcal{B}$ means that argument $\mathcal{A}$ attacks argument $\mathcal{B}$. Abstract argumentation frameworks can be concisely represented by directed graphs, where arguments are represented as nodes and edges model the attack relation.

Example 1 Consider the abstract argumentation framework $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ depicted in Figure 1. Here it is $\mathsf{Arg} = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \mathcal{A}_4, \mathcal{A}_5\}$ and $\rightarrow = \{(\mathcal{A}_1, \mathcal{A}_2), (\mathcal{A}_2, \mathcal{A}_1), (\mathcal{A}_2, \mathcal{A}_3), (\mathcal{A}_3, \mathcal{A}_4), (\mathcal{A}_4, \mathcal{A}_5), (\mathcal{A}_5, \mathcal{A}_4), (\mathcal{A}_5, \mathcal{A}_3)\}$.

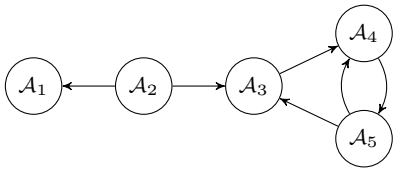

Fig. 1 A simple argumentation framework

Semantics are usually given to abstract argumentation frameworks by means of extensions [11]. An *extension* E of an argumentation framework $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ is a set of arguments $E \subseteq \mathsf{Arg}$ that gives some coherent view on the argumentation underlying AF.

In the literature [11,9] a wide variety of different types of semantics has been proposed. Here, we focus on

the grounded semantics [11] due to reasons of simplicity of presentation. Note that most works discussed in this overview do not rely on a specific semantics.

Definition 2 Let $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ be an argumentation framework.

1. An extension $E \subseteq \mathsf{Arg}$ is *conflict-free* iff there there are no $\mathcal{A}, \mathcal{B} \in E$ with $\mathcal{A} \rightarrow \mathcal{B}$.
2. An argument $\mathcal{A} \in \mathsf{Arg}$ is *acceptable* with respect to an extension $E \subseteq \mathsf{Arg}$ iff for every $\mathcal{B} \in \mathsf{Arg}$ with $\mathcal{B} \rightarrow \mathcal{A}$ there is $\mathcal{A}' \in E$ with $\mathcal{A}' \rightarrow \mathcal{B}$.
3. An extension $E \subseteq \mathsf{Arg}$ is *admissible* iff it is conflict-free and all $\mathcal{A} \in E$ are acceptable with respect to E.
4. An extension $E \subseteq \mathsf{Arg}$ is *complete* iff it is admissible and there is no $\mathcal{A} \in \mathsf{Arg} \setminus E$ which is acceptable with respect to E.
5. An extension $E \subseteq \mathsf{Arg}$ is *grounded* iff it is complete and E is minimal with respect to set inclusion.

The intuition behind admissibility is that an argument can only be accepted if there are no attackers that are accepted and if an argument is not accepted then there has to be an acceptable argument attacking it. The idea behind the completeness property is that all acceptable arguments should be accepted. The grounded extension is the minimal set of acceptable arguments and uniquely determined [11]. It can also easily be computed as follows: first, all arguments that have no attackers are added to an empty extension E and those arguments and all arguments that are attacked by one of these arguments are removed from the framework; then process is repeated; if one obtains a framework where there is no unattacked argument the remaining arguments are also removed.

Example 2 Consider again the argumentation framework AF in Figure 1. The grounded extension E_{gr} of AF is given by $E = \{\mathcal{A}_2\}$.

Abstract argumentation frameworks are arguably the most investigated formalism for argumentation and most works on strategic argumentation consider them as well. However, there are also formalisms for structured argumentation, such as deductive argumentation [7] and defeasible logic programming [14]. In structured argumentation, arguments are a set of (e. g. propositional) formulas (the support of an argument) that derive a certain conclusion (the claim of an argument). The attack relation between arguments is then derived from logical inconsistency. Although there are some works on strategic argumentation that work with structured approaches to argumentation, such as [34,37], we do not consider them here in depth to to lack of space.

3 Argumentation Dialogues and Games

The general setting of argumentation in multi-agent systems considers sets of agents that are engaged in a dialogue and exchange arguments. There are several different purposes of such a dialogue like negotiation, persuasion, information-seeking, inquiry, and deliberation, cf. [40]. A negotiation dialogue has the aim to distribute some given resources between the agents [16] while in a persuasion dialogue one agent aims at convincing the other agents of some beliefs [26,12]. In an information-seeking dialogue one agent aims at finding an answer by collecting arguments from other agents [39], while in an inquiry dialogue all agents seek to collaboratively find an answer to a question [38,8]. Finally, a deliberation dialogue is about jointly agreeing on a specific course of action [2,19].

Most works on argumentation dialogues are concerned with formalizing the interaction between agents, i. e., the locutions and the protocol. For example, in [8] an inquiry dialogue system is presented that allows agents to exchange structured arguments—built using *Defeasible Logic Programming* [14]—in order to collaboratively discover whether some claim can be accepted. In [8], Black and Hunter describe a protocol that prescribes legal orders of locutions that take into account relevance of replies to inquiries. The protocol consists of two sub processes, one on argument inquiry (how to build arguments using different agents' knowledge) and on warrant inquiry (how to relate arguments to each other in order to determine which argument can be accepted). Besides the formalization of the protocol they also give a simple implementation for the agents. A general discussion of argumentation protocols is given in [21].

Many of the above described types of argumentation dialogues offer the possibility of strategic argumentation. However, in most works on strategic argumentation the persuasion dialogue is used and we will also focus on this kind of dialogue in the following, see also [26] for a survey on persuasion dialogues that focuses more on the aspects of protocols and interaction than strategic behavior. The problem of strategic

argumentation in multi-agent systems can be best described with game-theoretical means. Agents engaging in a (persuasion) dialogue aim at establishing a certain goal. In general, this amounts to convincing the other agents that a certain statement is true. In the setting of abstract argumentation this usually amounts to showing that a certain argument (or one argument out of a set of arguments) should either be accepted or rejected by the grounded extension. Through strategic argumentation—i. e. forwarding only a specific subset of known arguments—agents try to reach this goal. For reasons of simplicity we consider only a simplified setting for strategic argumentation in multi-agent systems consisting of two agents, PRO and OPP. The goal of PRO is to establish a specific given argument $\mathcal{A}$ and the goal of OPP is to avoid this.

Example 3 Consider the abstract argumentation framework $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ depicted in Figure 2 and assume that PRO's goal is to establish that $\mathcal{A}_1$ is accepted. Note first that in the grounded extension of AF the argument $\mathcal{A}_1$ is not included and assume that OPP does not know the argument $\mathcal{A}_3$. Then PRO can act strategically by only putting forward arguments $\mathcal{A}_1$ and $\mathcal{A}_4$. Now, there is no way for OPP to disprove $\mathcal{A}_1$.

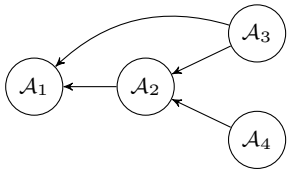

Fig. 2 The argumentation framework from Example 3

The scenario described in the example above is quite simple and so is the winning strategy for PRO: do not disclose arguments that may harm your own goal.

In [37] a classification of argumentation games has been proposed that brings certain complexities and opportunities for strategic argumentation. In particular, [37] discusses three different dimensions (or parameters) that constitute an argumentation game:

1. *Game protocol*: The exact way agents interact with each other constrains the opportunities for strategic argumentation very strictly. For example, a *direct game protocol*, which only allows a single step in the argumentation process and demands from each agent to bring forward a single set of arguments at once, does not allow for agents to react on other agents' moves. A standard *dialogue protocol*—where first one agent advances some arguments, then another agent reacts with some other arguments, etc. until no agent wants to advance further arguments—is a more dynamic setting with opportunities to react on what other agents bring forward and act appropriately.
2. *Awareness*: Whether or not an agent has background knowledge on other agents' beliefs influences its behavior. An *ignorant* agent, which only knows of the arguments itself is aware of but has no idea on what arguments other agents know of, is limited in its strategic capabilities. An *omniscient* agent, which knows what arguments other agents know of (and also if and what other agents believe that the first agents believes, etc.), has usually an advantage. It can simulate how other agents might react on moves and act accordingly.
3. *Goal types*: The way the goals of agents are organized also influences strategic argumentation. If an agent only has the goal to prove (or disprove) a single argument, its actions can focus on this particular task. If an agent aims at establishing a whole set of arguments (and maybe also to disprove another set) or maximize the number of arguments to be accepted from a given set, strategic argumentation has to be more sophisticated.

The examples above for the different dimensions are only corner cases that show how different instantiations of these dimensions may influence the opportunities for strategic argumentation. In between those examples there may be a whole space of different instantiations, each constraining the way strategic argumentation can be implemented. In particular, the dimension *awareness* can be instantiated by a series of different *opponent models* where beliefs one agent has about another can be captured. This might also take qualitative or quantitative uncertainty into account. We will have a particular look on opponent models in Section 6.

However, the dimensions listed above do not describe the setting of strategic argumentation completely. There are many further properties of argumentation games that are usually assumed to have a specific instantiation for reasons of simplicity. One such property, for example, is about the structure of the underlying ar-

gumentation framework and whether that is mutually agreed upon. In many works on strategic argumentation with abstract arguments such as [29,33] and almost all works on strategic argumentation with structured arguments such as [34] the attack relation between arguments is fixed or directly inferred from the underlying logic: if two arguments are put forward by possibly different agents, all agents agree if one argument attacks the other or not (even if an agent did not know the argument before). There are some works which do not make this assumption, see e. g. [15,16,22]. In particular, [15] discuss argumentation dialogues where the argumentation framework under consideration is not known with certainty. They use this framework to model argumentation in front of an audience and the goal is to persuade the audience rather than the opponent. The uncertainty of the framework then represents the uncertainty on the beliefs of the audience. In [15] strategies are discussed how to act in these settings. Furthermore, [16] deal with negotiation on offers. Arguments are exchanged and agents learn the attack relation of the opponent while negotiating. Finally, [22] use value-based argumentation frameworks [5] where the ordering of the values of the arguments (and thus the topology of the argumentation framework) is not fixed.

For the rest of this paper we focus on the setting where agents have a mutual agreement on whether an argument attacks another one or not. More specifically, we assume that there is a *universal* argumentation framework $\mathsf{AF} = (\mathsf{Arg}, \rightarrow)$ which contains all arguments relevant to a particular discourse (but parts of it maybe unknown to agents until some agent puts them forward).

Another property that may have an influence on the adopted strategies is the cost of the argumentation, cf. [34]. Costs can occur for an agent during argumentation for several different reasons:

- Costs in producing an argument: to construct an argument a reasoning process may be called that would take time and resources. For example, to produce a convincing argument that the shadows on the pictures of the moon-landing are indeed inconsistent, one could gather some reliable persons, fly to the moon an re-enact the original moon-landing. While the resulting argument would be a very strong one (given that it *could* be produced in this fashion), the costs in obtaining it are very high. Sometimes it is more beneficial to rely on simple arguments if the outcome of the dialogue is not so important.
- Costs of lengthy argumentation: argumentation may take a long time to reach a conclusion. In particular, when it comes to negotiation on goods it is sometimes beneficial to concede early in a discussion to avoid failing the whole dialogue [16].
- Costs incurred by information disclosure: every argument disclosed in a dialogue brings also new information for the opposing party. Information disclosed in this way may be to an agent's disadvantage in the long run. For example, consider the argument "the moon-landing did not take place as no living being can survive in space due to the Van Allen radiation" and assume that the agent who produced this argument is later engaged in a dialogue where he argues that the UFO landing really happened in Roswell in 1947. Then his own argument can be used against him as aliens could not have travelled space then (assuming aliens can be regarded as living beings).

For some discussion on including costs into the argumentation process see e. g. [34,16].

4 A Formal Model for Argumentation Games

In order to continue the discussion on strategic argumentation we will now introduce a very general formalization of argumentation games, see also [28,33] for some more concrete formalizations. First, we need the definition of a dialogue trace which is a sequence of moves in a dialogue.

Definition 3 A *dialogue trace* $M = (A_1, \ldots, A_n)$ is a sequence of sets of arguments $A_i \subseteq \mathsf{Arg}$.

A dialogue trace describes the history of a specific dialogue as it records which (sets of) arguments have been brought forward so far. Every dialogue trace $M = (A_1, \ldots, A_n)$ induces a view AF_M on the universal argumentation framework via $\mathsf{AF}_M = (A_1 \cup \ldots \cup A_n, \rightarrow \cap((A_1 \cup \ldots \cup A_n) \times (A_1 \cup \ldots \cup A_n))$ which is the argumentation framework both agents currently see as valid. Let $\mathcal{M}$ be the set of all dialogue traces. A *utility function* u is any function $u : \mathcal{M} \rightarrow \mathbb{R}$ that evaluates a dialogue trace M to a real value indicating its utility for the current agent (a larger value means a higher utility). An agent is characterized by its *belief state* $\mathcal{K}$ which contains the set of arguments he knows about and possibly its opponent model. Let $\mathbb{K}$ be the set of all possible belief states. Every agent has a *move function* $\mathsf{move} : \mathcal{M} \times \mathbb{K} \rightarrow 2^{\mathsf{Arg}}$ that returns the agent's

move, given the current dialogue trace and its belief state, and an *update function* $\mathsf{upd} : \mathcal{M} \times \mathbb{K} \to \mathbb{K}$ that updates an agent's belief state with new information from the current dialogue trace.

Definition 4 Let u be a utility function, $\mathcal{K}$ some belief state, move a move function, and upd an update function. Then $A = (u, \mathcal{K}, \mathsf{move}, \mathsf{upd})$ is called an *agent*.

As mentioned before, we constrain our attention to multi-agent systems with two agents PRO and OPP.

Definition 5 A protocol P is a function $P : \mathbb{N}^+ \to 2^{\{\mathsf{PRO},\mathsf{OPP}\}}$.

A protocol assigns to each round of an argumentation game the agents who are going to move. Examples of protocols are the *direct argumentation protocol* P_d defined as $P(0) = \{\mathsf{PRO}, \mathsf{OPP}\}$ and $P(i) = \emptyset$ for all $i > 0$ or the *round-robin protocol* P_r defined via $P(i) = \mathsf{PRO}$ for i even and $P(j) = \mathsf{OPP}$ for j odd, see also [37].

Definition 6 Let AF be an argumentation framework, P a protocol, and PRO and OPP two agents. Then $G = (\mathsf{AF}, P, \mathsf{PRO}, \mathsf{OPP})$ is called an *argumentation game*.

An argumentation game is played by iteratively calling the move functions of the agents in the way ascribed by the protocol. More specifically, the *induced dialogue trace* is defined as follows.

Definition 7 Let $G = (\mathsf{AF}, P, \mathsf{PRO}, \mathsf{OPP})$ be an argumentation game with $\mathsf{PRO} = (u_{\mathsf{PRO}}, \mathcal{K}^1_{\mathsf{PRO}}, \mathsf{move}_{\mathsf{PRO}}, \mathsf{upd}_{\mathsf{PRO}})$ and $\mathsf{OPP} = (u_{\mathsf{OPP}}, \mathcal{K}^1_{\mathsf{OPP}}, \mathsf{move}_{\mathsf{OPP}}, \mathsf{upd}_{\mathsf{OPP}})$. Then the *induced dialogue trace* $M_G = (A_1, \ldots, A_n)$ of G is defined as

1. $A_1 = \mathsf{move}_{P(1)}((), \mathcal{K}_{P(1)})$
2. $A_i = \mathsf{move}_{P(i)}((A_1, \ldots, A_{i-1}), \mathcal{K}^{i-1}_{P(i)})$ for all $i = 2, \ldots, n-2$
3. $A_{n-1} = A_n = \emptyset$

where $\mathcal{K}^i_A = \mathsf{upd}((A_1, \ldots, A_i), \mathcal{K}^{i-1}_A)$ for $i = 2, \ldots, n-2$ and $A \in \{\mathsf{PRO}, \mathsf{OPP}\}$.

The first item in the above definition states that the first move is made by the first player on the empty dialogue trace. The second item states that moves are made as described by the protocol. The final item describes the termination criterion of the game, i. e., the game ends when both agents consecutively make an empty move. Furthermore, the belief state of every agent has to be updated after every move. The final argumentation framework AF_{M_G} and its grounded extension E describe the outcome of the game. In particular, if $u_A(M_G) > 0$ then A is called a *winner* of the game for $A \in \{\mathsf{PRO}, \mathsf{OPP}\}$. Otherwise A is called a *loser* of the game.

Please note that the formalization above only roughly describes the common parts of most approaches to strategic argumentation but it will be sufficient for discussion in the remainder of this paper. For more elaborate formalization see the corresponding research works.

5 Strategic Argument Selection

The work [29] introduces *mechanism design* for argumentation games, see also [30]. *Mechanism design* deals with the question of whether strategic argumentation is beneficial at all in some settings and how to design an argumentation game and its protocol (i. e. its mechanism) so that strategic argumentation is useless.

The core notion here is *strategy-proofness*. Let $\mathsf{Arg}_{\mathsf{PRO}}$ be the set of arguments PRO knows about. A game $G = (\mathsf{AF}, P, \mathsf{PRO}, \mathsf{OPP})$ is called *strategy-proof* (for PRO) if under all variants of G where only the move function of PRO is modified, the *truthful strategy* $\mathsf{move}^t_{\mathsf{PRO}} = \mathsf{Arg}_{\mathsf{PRO}}$ yields maximal utility for PRO on M_G. This means that the *dominant* strategy for PRO is to truthfully report all arguments it knows of. Such games do not provide an opportunity for strategic exploitation and are thus preferred for application scenarios where strategic behavior should be avoided, such as medical applications. Furthermore, if a game is strategy-proof for all agents it is also computationally attractive as the protocol can always be implemented by a direct protocol, i. e., all agents report all their arguments in a single step. The research challenge in mechanism design for argumentation games is to find criteria or characterizations of strategy-proof games. These criteria may be topological criteria on the argumentation frameworks. For example, every argumentation framework without attacks leads (trivially) to a strategy-proof argumentation game. Other criteria can be about the utility functions of the agents. For example, [29] showed that if the goal of each agent is to maximize acceptance of the number of arguments from a given set and this set is conflict-free and contains no indirect attacks, then the corresponding game is strategy-proof. In [24] the investigation is extended to not only include grounded semantics but also preferred semantics, cf. [11].

There are a lot of cases where the results discussed above cannot be applied. Therefore, there are settings where strategic behavior is beneficial in order to reach a desired outcome of the argumentation. The question that arises is *how* to act strategically in a given setting. For example, consider the argumentation framework AF_0 depicted in Figure 3 and assume PRO wants to establish $\mathcal{A}_1$ and that PRO only knows of the arguments $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$. From PRO's perspective there is no reason to *not* put forward all his arguments as the grounded extension E of $\mathsf{AF}_{\{\mathcal{A}_1,\mathcal{A}_2,\mathcal{A}_3\}}$ contains $\mathcal{A}_1$, as desired. However, by putting forward $\mathcal{A}_3$ there is an opportunity for OPP to challenge $\mathcal{A}_1$ by putting forward $\mathcal{A}_4$. In that case, it would have been better for PRO to not disclose $\mathcal{A}_2$ as it may be used (if e. g. defended by $\mathcal{A}_4$) to defeat $\mathcal{A}_1$. More generally, if no further information is available—i. e. if PRO has no beliefs on what OPP believes—then the best strategy for PRO is to not disclose potentially harmful arguments such as $\mathcal{A}_2$. This strategy has been called *overcautious strategy* in [37] and can be used as a heuristic for direct argumentation protocols. In dialectical protocols such as the round-robin protocol it may be necessary to relax the strategy a bit, see e. g. the argumentation framework AF_1 in Figure 4. If PRO starts by putting forward $\mathcal{A}_1$ and OPP reacts with $\mathcal{A}_5$ then it would be beneficial for PRO to react with $\mathcal{A}_4$, even if $\mathcal{A}_4$ can also be used to defeat $\mathcal{A}_1$ along the path $\mathcal{A}_4, \mathcal{A}_3, \mathcal{A}_2, \mathcal{A}_1$. If an agent has no further information on what the other agent knows the above outlined strategy is a baseline approach for strategic behavior in argumentation games. If we allow agents to be aware of other agents' beliefs more opportunities for strategic argumentation arise. In the following section we have a specific look at opponent models that exactly serve this purpose.

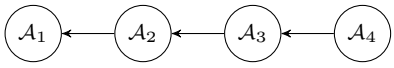

Fig. 3 The argumentation framework AF_0

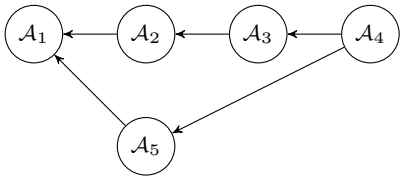

Fig. 4 The argumentation framework AF_1

6 Opponent Models

[23] introduced a recursive opponent model for strategic argumentation. This opponent model can be formalized as a tuple $E_0 = (B_0, E_1)$ where B_0 is a set of arguments and $E_1 = (B_1, E_2)$ is itself and opponent model. Assume that E_0 is the opponent model agent PRO has about OPP, i. e., it is some component in PRO's belief state $\mathcal{K}_{\mathsf{PRO}}$. Then B_0 is the set of arguments PRO believes OPP to know about and B_1 is the set of arguments that PRO believes that OPP believes that PRO knows about, etc. By employing a variant of the Maxmin-algorithm [10] this model can be used for strategic argumentation: when PRO has to execute a move he first simulates how OPP would react given B_0 (which is itself dependent on how PRO would react given B_1) and then selects the move that maximizes PRO's utility given the reaction of OPP. This model has been extended by [33] with qualitative uncertainty on both the opponent model and the set of arguments. For example, instead of an opponent model of the form $E_0 = (B_0, E_1)$ one considers an opponent model $E_0 = (B_0, P_0)$ where P_0 is a probability distribution over opponent models (which themselves contain probability distributions over opponent models, etc.).

Usually, having an opponent model is beneficial for strategic argumentation as it enables an agent to make a better informed decision. However, as in many multiplayer games investigated with game-theoretical means also strategic argumentation with opponent models may suffer from the *paradox of omniscience*, cf. [25]. Consider the following example.

Example 4 Imagine the game of "chicken": two drivers A and B are each sitting in a car and driving towards each other. Each driver may either drive straight or veer. If both drivers drive straight they crash and they will both die. If either one of them drives straight and the other veers the latter one is the loser of the game and the former is the winner. If both drivers veer both lose. This game can be represented as the argumentation framework depicted in Figure 5. A driver can

only veer or drive straight making the corresponding arguments mutually exclusive. Furthermore, both drivers cannot drive straight at the same time as this results in a crash. The utility function of driver A is defined such that the outcome $\{S_A, V_B\}$ is the most preferred one, $\{V_A, V_B\}$ the second most preferred one, $\{V_A, S_B\}$ the third, and $\{S_A, S_B\}$ the worst one. The utility function of B is defined analogously. Finally, A only knows of arguments V_A and S_A and B only of V_B and S_B.

If one considers a direct argumentation protocol without opponent models, the best move for both agents is to move with V_A and V_B, respectively. Furthermore, even if both agents have a complete opponent model (e. g. every agent knows that every agent knows every argument) the best option for both agents is to veer. However, so far we have only considered a model of the opponent that describes what the opponent *believes* and not how he is going to *act*. Assume that driver A is really omniscient, i. e., he has a perfect opponent model and also knows how driver B will act in the game of "chicken" (i. e. A knows whether B will drive straight or veer) and assume that B only knows that A is omniscient. Now, even in the direct argumentation protocol, B can put forward S_B (driving straight) without any worries as B knows that A knows his decision beforehand and must therefore veer (putting forward V_A). In this case, the more sophisticated opponent model of A is a disadvantage.

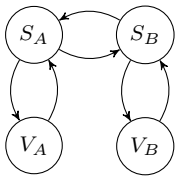

Fig. 5 The argumentation framework from Example 4 representing the game of "chicken" with arguments S_A (driver A drives straight), S_B (driver B drives straight), V_A (driver A veers), V_B (driver B veers).

In the current state of the art of opponent modeling for strategic argumentation only opponent models describing what another agent believes are used. In particular, it is usually assumed that agents follow the same type of strategy but with different utility functions. The strategy followed by player B in the above example is a meta-strategy that first analyzes the strategy of A and then selects a strategy for himself. These kinds of meta-strategies have not been investigated for strategic argumentation so far.

One question that arises when considering opponent models as a means for capturing the beliefs an agent has about another agent, is how did the agent obtain these beliefs? In the setting of [23,33] these beliefs are assumed to be given which is an unrealistic assumption in most application settings. However, one way of acquiring these beliefs is by experience and learning from previous argumentation dialogues. The setting of strategic argumentation we considered so far is a one-shot scenario: two agents argue about a certain argument and after the dialogue is finished, the protocol ends. However, agents are usually engaged in a series of dialogues, either about the same argument with different agents, with the same agent about different arguments, or combinations of those. In this more general setting, the *epistemic* component of arguments can be exploited in order to learn the behavior of agents. More specifically, arguments are no mere pieces of information that attack each other but can be in other relationships as well, as can also be formalized using structured approaches to argumentation [14,7]. For example, arguments may support each other and, in particular, the awareness of a specific argument may imply the awareness of another argument. Consider the example from the introduction about conspiracy theories regarding the first moon-landing. After Anna presents her first argument about the inconsistent shadows in the pictures, Bob might come to believe that Anna has done some fair reading on the moon-landing and its conspiracy theories. So he might already believe (up to a certain degree) that Anna will also have some arguments regarding e. g. the fluttering flag.

The relationships of arguments with respect to their *mutual appearance* can be learned by engaging in multiple dialogues and observing these co-occurrences multiple times. In [17] the authors exactly follow this approach and learn opponent models from experience. A *relationship graph* records co-occurrences of arguments brought forward by other agents and this graph is used to predict of what other arguments a particular agents knows given a partial observation of that agent's behavior.

7 Further Works

In this overview paper we focused on strategic argument selection with respect to argumentation games with grounded semantics. There are other works which discuss strategic aspects in argumentation but do not entirely fit this framework. We will have a look at some of them in this section.

The work [32] deals with merging labelings (which are a generalization of extensions). In their setting, the argumentation framework AF is fixed and known to all agents. However, the agents may disagree on what labeling/extension to use to evaluate AF. Recall that for other semantics than grounded semantics, the labeling/extension conforming to this semantics may not be uniquely determined, cf. [3]. For example, in an argumentation framework consisting of two arguments $\mathcal{A}_1$ and $\mathcal{A}_2$ with a mutual attack between them ($\mathcal{A}_1 \to \mathcal{A}_2$ and $\mathcal{A}_2 \to \mathcal{A}_1$) there are two *preferred* extensions E_1 and E_2 with $E_1 = \{\mathcal{A}_1\}$ and $E_2 = \{\mathcal{A}_2\}$. In such a setting, different agents may adopt different labelings/extensions for evaluation and [32] deal with the question of how to merge this set of labelings/extensions into a single one that can be used for collective evaluation. This problem is closely related to the problem of judgement aggregation [1] and, therefore, also exploitable by strategic manipulation. Agents can lie about their labeling/extension in order to manipulate the merging process and the final outcome.

The paper [18] discusses strategic behavior for argumentation in social contexts. In that paper the term *strategy* has a slightly different meaning than we used here. The work [18] describes a multi-agent setting with social obligations and presents a framework for resolving conflicts of obligations through negotiation. Strategies are then used to deal with failed negotiations such as by demanding compensation for not fulfilling an obligation or by incorporating threats or promises into the argumentation process.

In [35] the authors use defeasible logic as the means to represent beliefs and as the building blocks for arguments. They consider dialogues of agents exchanging formulas and use game trees to analyze and predict expected outcomes. These predictions can then be used to guide argument selection.

8 Discussion

This paper gave a brief overview of the field of strategic argumentation in multi-agent systems. We discussed general properties of argumentation dialogues and approaches for strategic exploitation.

One challenge of research in strategic argumentation concerns its evaluation. Usually, research in computational models of argumentation is evaluated analytically by proving certain desirable properties or relating the work to other fields such as other approaches to non-monontonic reasoning. However, the analysis of approaches to argumentation in multi-agent systems becomes complex very fast, in particular, if non-trivial examples of dialogues are studied. Although most researchers in strategic argumentation come from knowledge representation research, many have adopted now empirical evaluation methods to show the feasibility of their approaches, as is also common in other subfields of multi-agent systems research. Some examples of works employing empirical evaluation (mostly on artificially generated argumentation frameworks and argumentation games) are [18,16,33]. The *Tweety libraries for logical aspects of artificial intelligence and knowledge representation*[1] also contain an evaluation framework for strategic argument selection as it has been discussed in this paper.

In this overview paper we only tackled the issue of strategic argumentation on abstract argumentation frameworks. When considering structured argumentation frameworks such as ASPIC [27], *Defeasible Logic Programming* [14], or *deductive argumentation* [7] further issues relating to strategical behavior arise. In structured argumentation frameworks arguments are built by combining smaller logical elements such as rules and facts. The attack relation in these frameworks is then usually derived by using logical contradiction, e. g,. an argument claiming a proposition a by using the rule $b \to a$ and the fact b attacks an argument claiming c which uses the rules $d \to \neg a$ and $\neg a \to c$ and the fact d. When agents exchange rules, facts, and arguments the possibility arises that these elements can be combined to new arguments that have been unknown before. There are only few works on strategic issues for structured argumentation but some discussion can be found in [37].

[1] `http://www.mthimm.de/projects/tweety/`

Approaches to strategic argumentation can be used e. g. for decision-support tools for legal reasoning [35, 34, 4, 36] or for autonomous negotiation agents [18, 16]. Furthermore, research in strategic argumentation also helps in understanding how humans act strategically in dialogues and how their behavior can be predicted. The research field is still quite young and there are a lot of opportunities to advance it further.

Acknowledgements I thank Tjitze Rienstra, Nir Oren, Tony Hunter, Gabriele Kern-Isberner, and Alejandro García for valuable discussions on the topic of computational models of argumentation in general and strategic argumentation in particular.

References

1. Arrow, K.J.: Social Choice and Individual Values. Wiley (1951)
2. Atkinson, K., Bench-Capon, T.J.M., McBurney, P.: A Dialogue Game Protocol for Multi-Agent Argument over Proposals for Action. In: I. Rahwan, P. Moraitis, C. Reed (eds.) Proceedings of the First International Workshop on Argumentation in Multi-Agent Systems (ArgMAS'04), pp. 149–161. Springer (2004)
3. Baroni, P., Caminada, M., Giacomin, M.: An Introduction to Argumentation Semantics. The Knowledge Engineering Review **26**(4), 365–410 (2011)
4. Beierle, C., Freund, B., Kern-Isberner, G., Thimm, M.: Using Defeasible Logic Programming for Argumentation-Based Decision Support in Private Law. In: P. Baroni, F. Cerutti, M. Giacomin, G.R. Simari (eds.) Proceedings of the Third International Conference on Computational Models of Argument (COMMA'10), pp. 87–98. IOS Press (2010)
5. Bench-Capon, T.J.M.: Persuasion in Practical Argument Using Value Based Argumentation Frameworks. Journal of Logic and Computation **13**(3), 429–448 (2003)
6. Bench-Capon, T.J.M., Dunne, P.E.: Argumentation in artificial intelligence. Artificial Intelligence **171**(10–15), 619–641 (2007)
7. Besnard, P., Hunter, A.: Elements of Argumentation. The MIT Press (2008)
8. Black, E., Hunter, A.: An Inquiry Dialogue System. Autonomous Agents and Multi-Agent Systems **19**(2), 173–209 (2009)
9. Caminada, M.: Semi-Stable Semantics. In: P. Dunne, T. Bench-Capon (eds.) Proceedings of the First International Conference on Computational Models of Argument (COMMA'06), pp. 121–130. IOS Press (2006)
10. Carmel, D., Markovitch, S.: Learning and Using Opponent Models in Adversary Search. Technical Report CIS9609, Technion (1996)
11. Dung, P.M.: 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)
12. Fan, X., Toni, F.: Mechanism Design for Argumentation-based Persuasion. In: B. Verheij, S. Szeider, S. Woltran (eds.) Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA'12), pp. 322–333. IOS Press (2012)
13. Fermé, E.L., Gabbay, D.M., Simari, G.R. (eds.): Trends in Belief Revision and Argumentation Dynamics. College Publications (2014)
14. Garcia, A., Simari, G.R.: Defeasible Logic Programming: An Argumentative Approach. Theory and Practice of Logic Programming **4**(1–2), 95–138 (2004)
15. Grossi, D., van der Hoek, W.: Audience-Based Uncertainty in Abstract Argument Games. In: F. Rossi (ed.) Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13), pp. 143–149 (2013)
16. Hadidi, N., Dimopoulos, Y., Moraitis, P.: Tactics and Concessions for Argumentation-based Negotiation. In: B. Verheij, S. Szeider, S. Woltran (eds.) Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA'12), pp. 285–296. IOS Press (2012)
17. Hadjinikolis, C., Siantos, Y., Modgil, S., Black, E., McBurney, P.: Opponent Modelling in Persuasion Dialogues. In: F. Rossi (ed.) Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13), pp. 164–170 (2013)
18. Karunatillake, N.C., Jennings, N.R., Rahwan, I., McBurney, P.: Dialogue Games that Agents Play within a Society. Artificial Intelligence **173**(9-10), 935–981 (2009)
19. Kok, E.M., Meyer, J.J.C., Prakken, H., Vreeswijk, G.A.W.: A Formal Argumentation Framework for Deliberation Dialogues. In: P. McBurney, I. Rahwan, S. Parsons (eds.) Proceedings of the Seventh International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2010), pp. 73–90 (2010)
20. McBurney, P., Parsons, S., Rahwan, I. (eds.): Proceedings of the Eighth International Workshop on Argumentation in Multi-Agent Systems (ArgMAS'12), *Lecture Notes in Computer Science*, vol. 7543. Springer (2012)
21. McBurney, P., Parsons, S., Wooldridge, M.: Desiderata for Agent Argumentation Protocols. In: M. Gini, T. Ishida, C. Castelfranchi, W.L. Johnson (eds.) Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems (AAMAS'02) (2002)
22. Oren, N., Atkinson, K., Li, H.: Group persuasion through uncertain audience modelling. In: B. Verheij, S. Szeider, S. Woltran (eds.) Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA'12), pp. 350–357. IOS Press (2012)
23. Oren, N., Norman, T.J.: Arguing using Opponent Models. In: P. McBurney, I. Rahwan, S. Parsons, N. Maudet (eds.) Proceedings of the Sixth International Workshop on Argumentation in Multi-Agent Systems (ArgMAS'09), pp. 160–174. Springer (2010)
24. Pan, S., Larson, K., Rahwan, I.: Argumentation Mechanism Design for Preferred Semantics. In: P. Baroni, F. Cerutti, M. Giacomin, G.R. Simari (eds.) Proceedings of the Third International Conference on Computational Models of Argument (COMMA'10), pp. 403–414. IOS Press (2010)

25. Poundstone, W.: Labyrinths of Reason: Paradox, Puzzles and the Frailty of Knowledge. Penguin Books (1988)
26. Prakken, H.: Formal Systems for Persuasion Dialogue. The Knowledge Engineering Review **21**, 163–188 (2006)
27. Prakken, H.: An Abstract Framework for Argumentation with Structured Arguments. Argument and Computation **1**(2), 93–124 (2010)
28. Procaccia, A.D., Rosenschein, J.S.: Extensive-Form Argumentation Games. In: Proceedings of the Third European Workshop on Multi-Agent Systems (EUMAS'05), pp. 312–322 (2005)
29. Rahwan, I., Larson, K.: Mechanism Design for Abstract Argumentation. In: L. Padgham, D. Parkes (eds.) Proceedings of Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS'08), pp. 1031–1038 (2008)
30. Rahwan, I., Larson, K.: Argumentation and Game Theory. In: I. Rahwan, G.R. Simari (eds.) Argumentation in Artificial Intelligence, pp. 321–339. Springer (2009)
31. Rahwan, I., Larson, K., Tohmé, F.: A Characterisation of Strategy-Proofness for Grounded Argumentation Semantics. In: C. Boutilier (ed.) Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09), pp. 251–256 (2009)
32. Rahwan, I., Tohmé, F.: Collective Argument Evaluation as Judgement Aggregation. In: W. van der Hoek, G.A. Kaminka, Y. Lespérance, M. Luck, S. Sen (eds.) Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010), pp. 417–424 (2010)
33. Rienstra, T., Thimm, M., Oren, N.: Opponent Models with Uncertainty for Strategic Argumentation. In: F. Rossi (ed.) Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13), pp. 332–338 (2013)
34. Riveret, R., Prakken, H., Rotolo, A., Sartor, G.: Heuristics in Argumentation: A Game-Theoretical Investigation. In: P. Besnard, S. Doutre, A. Hunter (eds.) Proceedings of the Second International Conference on Computational Models of Argument (COMMA'08), pp. 324–335. IOS Press (2008)
35. Roth, B., Riveret, R., Rotolo, A., Governatori, G.: Strategic Argumentation: A Game Theoretical Investigation. In: A. Gardner, R. Winkels (eds.) Proceedings of the Eleventh International Conference on Artificial Intelligence and Law (ICAIL'07), pp. 81–90. ACM Press (2007)
36. Thang, P.M., Dung, P.M., Hung, N.D.: Towards Argument-based Foundation for Sceptical and Credulous Dialogue Games. In: B. Verheij, S. Szeider, S. Woltran (eds.) Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA'12), pp. 398–409. IOS Press (2012)
37. Thimm, M., Garcia, A.J.: Classification and Strategical Issues of Argumentation Games on Structured Argumentation Frameworks. In: W. van der Hoek, G.A. Kaminka, Y. Lespérance, M. Luck, S. Sen (eds.) Proceedings of the Ninth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'10), pp. 1247–1254 (2010)
38. Thimm, M., Garcia, A.J., Kern-Isberner, G., Simari, G.R.: Using Collaborations for Distributed Argumentation with Defeasible Logic Programming. In: M. Pagnucco, M. Thielscher (eds.) Proceedings of the Twelfth International Workshop on Non-Monotonic Reasoning (NMR'08), pp. 179–188. University of New South Wales, Technical Report No. UNSW-CSE-TR-0819 (2008)
39. Thimm, M., Kern-Isberner, G.: A Distributed Argumentation Framework using Defeasible Logic Programming. In: P. Besnard, S. Doutre, A. Hunter (eds.) Proceedings of the Second International Conference on Computational Models of Argument (COMMA'08), pp. 381–392. IOS Press (2008)
40. Walton, D.N., Krabbe, E.C.W.: Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. SUNY Press (1995)

Matthias Thimm received his PhD from the University of Dortmund in 2011 on the topic of probabilistic reasoning with inconsistencies. Since 2011 he is a senior researcher at the Institute for Web Science and Technologies at the University of Koblenz-Landau. His research interests include computational models of argumentation, inconsistency resolution, and distributed approaches to knowledge handling, in particular with respect to linked data.