On the Interplay between Games, Argumentation and Dialogues

Xiuyi Fan
Nanyang Technological University, Singapore
xyfan@ntu.edu.sg

Francesca Toni
Imperial College London, United Kingdom
ft@imperial.ac.uk

ABSTRACT

Game theory, argumentation and dialogues all address problems concerning inter-agent interaction, but from different perspectives. In this paper, we contribute to the study of the interplay between these fields. In particular, we show that by mapping games in normal form into structured argumentation, computing *dominant solutions* and *Nash equilibria* is equivalent to computing *admissible sets of arguments*. Moreover, when agents lack complete information, computing dominant solutions/Nash equilibria is equivalent to constructing *successful (argumentation-based) dialogues*. Finally, we study agents' behaviour in these dialogues in reverse game-theoretic terms and show that, using specific notions of utility, agents engaged in (argumentation-based) dialogues are guaranteed to be *truthful and disclose relevant information*, and thus can converge to dominant solutions/Nash equilibria of the original games even under incomplete information.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence-multiagent systems, coherence and coordination

Keywords

Argumentation; Dialogues; Games

1. INTRODUCTION

Argumentation (see [3, 4, 37, 28] for overviews) provides means to reason with incomplete and inconsistent information, e.g. when this is held by collaborative or competing agents. Argumentation-based dialogues (e.g. see [2, 31, 32, 30, 6, 26, 39, 21]) study dialectical interactions amongst agents. Game theory (see e.g. [42] for an overview) studies agents' strategic behaviour. The interplay between these three fields brings interesting research opportunities, including addressing the following questions:

- **Q1:** Could argumentation be used to solve game-theoretic problems? If so, what are suitable representations, semantics and interaction models?
- **Q2:** Could game theory be used to analyse agents' behaviour in argumentation? If so, how to formulate behaviour and how to ensure "good" behaviour?

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

Both questions have been partially looked at before, for example, [9] studied how the stable marriage problem can be modelled using abstract argumentation (AA); [36, 34, 29] discussed conditions under which direct mechanisms based on grounded and preferred semantics for AA [9] are strategy-proof; [16, 25] analysed the computational complexity of behaving strategically in argumentation dialogues. We look at these questions in the context of *Assumption-based Argumentation (ABA)*, a form of structured argumentation (see [10, 41] for overviews). ABA is an instance of AA [11] and it admits AA as an instance [40]. ABA is well-studied with readily usable results, including a dialectical counterpart (referred to as *ABA dialogues* [15]) and soundness results connecting argumentation and dialogues [15]. ABA affords a finer level of granularity than AA, allowing, in particular, multiple agents to construct arguments jointly in dialogues, while assessing their acceptability as in AA. At the same time, ABA is an instance of other structured argumentation frameworks (such as ASPIC+ [27]) with just enough constructs to support our analysis in this paper (for example, preferences over rules, accommodated in ASPIC+ but not directly in ABA, are not required in our analysis).

For both **Q1** and **Q2**, we consider two agents α and β, such that:

1. each agent $x \in \{\alpha, \beta\}$ has a set of *strategies* Σ_x; strategies are *public*, namely each agent is aware of the set of strategies of the other;
2. each agent x is equipped with a *payoff function* π_x, namely a mapping from pairs of strategies to real numbers such that, for $(\sigma_\alpha, \sigma_\beta) \in \Sigma_\alpha \times \Sigma_\beta$, $\pi_x(\sigma_\alpha, \sigma_\beta)$ denotes the payoff value of $(\sigma_\alpha, \sigma_\beta)$ for x; the payoff functions are *private*, namely each agent is only aware of its own payoff function;
3. a *game* between α, β is a tuple $\langle \Sigma_\alpha, \Sigma_\beta, \pi_\alpha, \pi_\beta \rangle$.

We study two well-known solution concepts for these types of games: dominant solutions and Nash equilibria. A pair of strategies ($\sigma^\star_\alpha$, $\sigma^\star_\beta$) is a *dominant solution* iff for all pairs of strategies (σ_α, σ_β) it holds that

$$\pi_\alpha(\sigma^\star_\alpha, \sigma^\star_\beta) \geq \pi_\alpha(\sigma_\alpha, \sigma_\beta) \text{ and } \pi_\beta(\sigma^\star_\alpha, \sigma^\star_\beta) \geq \pi_\beta(\sigma_\alpha, \sigma_\beta).$$

Also, ($\sigma^\star_\alpha$,$\sigma^\star_\beta$) is a *Nash equilibrium* iff for all (σ_α, σ_β) it holds that

$$\pi_\alpha(\sigma^\star_\alpha, \sigma^\star_\beta) \geq \pi_\alpha(\sigma_\alpha, \sigma^\star_\beta) \text{ and } \pi_\beta(\sigma^\star_\alpha, \sigma^\star_\beta) \geq \pi_\beta(\sigma^\star_\alpha, \sigma_\beta).$$

In words, dominant solutions are strategy pairs which give the highest payoffs to both agents, whereas Nash equilibria are such that given the strategy adopted by the other agent, neither agent could do strictly better by adopting another strategy.

We illustrate games and solution concepts with a simple example, used throughout the paper.

		Wife (β)	
		r	**w**
Husband (α)	**m**	2,1	0,0
	f	0,0	1,2

Table 1: Dinner Party Game payoff matrix.

EXAMPLE 1.1. *(***Dinner Party Game***) Two agents ("husband" and "wife") want to organise a dinner party. They decide to split tasks: the husband will buy either meat (m) or fish (f) for the main course; the wife will buy either red wine (r) or white wine (w) as accompanying beverage. The husband (α) has strategies $\Sigma_\alpha = \{m, f\}$; the wife (β) has strategies $\Sigma_\beta = \{r, w\}$. The payoffs are given in Table 1, e.g. $\pi_\alpha(m, r) = 2, \pi_\beta(m, r) = 1$. Both (m, r) and (f, w) are Nash equilibria; there is no dominant solution for this game.*

Since we are assuming that payoffs are private (and thus agents have incomplete information), dominant solutions/Nash equilibria cannot be computed without some information exchange. In our model, agents use dialogues to support this exchange. As an illustration, for Example 1.1, agents can identify the Nash equilibrium (m, r) using the following dialogue (expressed in natural language here and formalised as an ABA dialogue later in Example 4.1):

Husband: meat and red wine make a good combination.
Husband: I assume you are OK with this choice.
Husband: Please let me know if you disagree.
Wife: It is not good if you prefer fish and red wine.
Husband: No, I don't.
Husband: And you don't prefer meat and white wine either?
Wife: No, I don't.

In this dialogue, the agents jointly identify a good combination of food and wine (a Nash equilibrium) by sharing (and disclosing) information, but no more than strictly necessary (e.g. the wife does not disclose her payoff for (m, r) or for any other strategy pair).

In this paper, we address **Q1** by using ABA dialogues to compute, in a distributed manner, dominant solutions and Nash equilibria for the kinds of games we consider (Section 4), based upon ABA formulations of games such that the solution concepts correspond to admissibility of arguments (Section 3). Moreover, we address **Q2** by modelling ABA dialogues themselves as games and studying dialectical strategies as agents' game strategies (Section 5). To ensure truthfulness and disclosure of relevant information from agents, we use reverse game theory (also known as mechanism design, see e.g. [19]) and define specific notions of utility to ensure truthfulness and disclosure in dialogue sequences (Section 6). As a result, we prove that agents engaged in ABA dialogues can converge to dominant solutions/Nash equilibria of the original games even under incomplete information.

2. BACKGROUND

Assumption-based Argumentation (ABA) *frameworks* [41] are tuples $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$ where

- $\langle \mathcal{L}, \mathcal{R} \rangle$ is a deductive system, with $\mathcal{L}$ the *language* and $\mathcal{R}$ a set of *rules* of the form $s_0 \leftarrow s_1, \ldots, s_m (m \geq 0, s_i \in \mathcal{L})$;
- $\mathcal{A} \subseteq \mathcal{L}$ is a (non-empty) set, whose elements are *assumptions*;
- $\mathcal{C}$ is a total mapping from $\mathcal{A}$ into $2^{\mathcal{L}} - \{\{\}\}$, where each $s \in \mathcal{C}(a)$ is a *contrary* of a, for $a \in \mathcal{A}$.

Given a rule $\rho = s_0 \leftarrow s_1, \ldots, s_m$, s_0 is referred to as the *head* (denoted $Head(\rho) = s_0$) and $s_1, \ldots, s_m$ as the *body* (denoted $Body(\rho) = \{s_1, \ldots, s_m\}$). All ABA frameworks in this paper are *flat*, namely such that no assumption is the head of a rule.

In ABA, *arguments* are deductions of claims using rules and supported by sets of assumptions, and *attacks* are directed at the assumptions in the support of arguments. Following [10]:

- *an argument for (claim) $s \in \mathcal{L}$ supported by $A \subseteq \mathcal{A}$* (denoted $A \vdash s$) is a finite tree with nodes labelled by sentences in $\mathcal{L}$ or by τ[1], the root labelled by s, leaves either τ or assumptions in A, and non-leaves s' with, as children, the elements of the body of some rule ρ with head s' (we say that ρ is *in* $A \vdash s$);
- *an argument* $A_1 \vdash s_1$ attacks an argument $A_2 \vdash s_2$ iff s_1 is a contrary of some assumption in A_2.

A *set of arguments* Δ *is admissible* iff Δ is *conflict-free* (i.e. no argument in Δ attacks any argument in Δ) and all arguments attacking some argument in Δ are counter attacked by arguments in Δ; an *argument is admissible* iff it belongs to an admissible set of arguments; a *sentence is admissible* iff it is the claim of an admissible argument.

Two ABA frameworks can be "merged" to form a single *joint framework* [13]: given $AF = \langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$, $AF' = \langle \mathcal{L}, \mathcal{R}', \mathcal{A}', \mathcal{C}' \rangle$, the joint framework (of AF and AF') is $AF_J = \langle \mathcal{L}, \mathcal{R} \cup \mathcal{R}', \mathcal{A} \cup \mathcal{A}', \mathcal{C}_J \rangle$, with $\mathcal{C}_J(a) = \mathcal{C}(a) \cup \mathcal{C}'(a)$, for all $a \in \mathcal{A} \cup \mathcal{A}'$.

ABA-dialogues [12, 15] are conducted between two agents, say α and β, that can be thought of as being equipped with ABA frameworks $\langle \mathcal{L}, \mathcal{R}_\alpha, \mathcal{A}_\alpha, \mathcal{C}_\alpha \rangle$ and $\langle \mathcal{L}, \mathcal{R}_\beta, \mathcal{A}_\beta, \mathcal{C}_\beta \rangle$ respectively, sharing a common language $\mathcal{L}$. An ABA-dialogue is made of *utterances* (from agent $x \in \{\alpha, \beta\}$) of the form $\langle x, y, T, C, ID \rangle$ (for $y \in \{\alpha, \beta\}, x \neq y$) where: C (the *content*) is one of: $claim(s)$ for some $s \in \mathcal{L}$, $rl(s_0 \leftarrow s_1, \ldots, s_m)$ for some $s_0, \ldots, s_m \in \mathcal{L}$, $asm(a)$ for some $a \in \mathcal{L}$, $ctr(a, s)$ for some $a, s \in \mathcal{L}$, a *pass* sentence $\pi \notin \mathcal{L}$; $ID \in \mathbb{N}$ (the *identifier*); $T \in \mathbb{N} \cup \{0\}$ (the *target*) such that $T < ID$. In the remainder, $\mathcal{U}^x$ will denote the set of all possible utterances from agent x.

A *dialogue* $\mathcal{D}^x_y(\chi)$ *(between x and y, $x, y \in \{\alpha, \beta\}$, $x \neq y$, for $\chi \in \mathcal{L}$)*, is a sequence $\langle u_1, \ldots, u_n \rangle$, $n \geq 0$, where each u_l, $l = 1, \ldots, n$, is an utterance, and: $u_1 = \langle x, y, \_, \_, \_ \rangle$;[2] the content of u_l is $claim(\chi)$ iff $l = 1$; the target of pass and claim utterances is 0; the target of regular utterances is not 0; the identifier of an utterance represents the position of the utterance in a dialogue, and the target of a non-pass, non-claim utterance is the identifier of some earlier utterance.

The *framework drawn from dialogue* $\delta = \langle u_1, \ldots, u_n \rangle$ is $F_\delta = \langle \mathcal{L}, \mathcal{R}_\delta, \mathcal{A}_\delta, \mathcal{C}_\delta \rangle$ where

- $\mathcal{R}_\delta = \{\rho | rl(\rho)$ is the content of some u_i in $\delta\}$;
- $\mathcal{A}_\delta = \{a | asm(a)$ is the content of some u_i in $\delta\}$;
- $\mathcal{C}_\delta(a) = \{s | ctr(a, s)$ is the content of some u_i in $\delta\}$.

Restrictions can be imposed on dialogues so that they fulfil desirable properties, and in particular that P1) the framework drawn from them is a flat ABA framework , and P2) utterances are related to their target utterances, where $u_j = \langle \_, \_, T, C_j, \_ \rangle$ is *related to* $u_i = \langle \_, \_, \_, C_i, ID \rangle$ iff $T = ID$ and one of the following holds:

- $C_j = rl(\rho_j)$, $Head(\rho_j) = s$ and either $C_i = rl(\rho_i)$ with $s \in Body(\rho_i)$, or $C_i = ctr(\_, s)$, or $C_i = claim(s)$;
- $C_j = asm(a)$ and either $C_i = rl(\rho)$ with $a \in Body(\rho)$, or $C_i = ctr(\_, a)$, or $C_i = claim(a)$;
- $C_j = ctr(a, \_)$ and $C_i = asm(a)$.

Properties P1) and P2) above can be enforced using *legal-move functions*, which are mappings from dialogues to sets of utterances such that there is no repeated utterance to the same target in a dialogue *compatible* with a legal-move function. We assume that dialogues compatible with legal-move functions defined later in the paper also satisfy both P1 and P2.

ABA dialogues can be used to check admissibility of claims: an ABA dialogue is deemed *successful* iff its claim is admissible

[1] $\tau \notin \mathcal{L}$ represents "true" and stands for the empty body of rules.
[2] Throughout, _ stands for an anonymous variable as in Prolog.

in the ABA framework drawn from the dialogue [15].[3] *Strategy-move* functions [13] can be used to model agents' behaviour, towards fulfillment of their aims, in ABA dialogues. A strategy-move function for agent x is a mapping ϕ from dialogues and legal-move functions to $\mathcal{U}^x$ such that $\phi(\delta, \lambda) \subseteq \lambda(\delta)$. Given a dialogue $\delta = \langle u_1, \ldots, u_n \rangle$ between agents x, y compatible with a legal-move function λ and a strategy-move function ϕ for x, if for all utterances u_m made by x, $u_m \in \phi(\langle u_1, \ldots, u_{m-1} \rangle, \lambda)$, then x *uses* ϕ in δ. If x and y both use ϕ, then δ is *constructed with* ϕ. The strategy-move function we use in this work is the *thorough strategy-move* function, ϕ_h. Informally, a dialogue constructed with ϕ_h contains all information that is relevant to the claim from both agents. Dialogues constructed with ϕ_h have the desirable property that admissible arguments obtained in the dialogue are admissible in the joint ABA framework of the two agents, see Theorem 1 in [13]. It is important to note that the thorough strategy-move function does not force agents to utter all knowledge but only information relevant to the claim.

Mechanism Design (e.g. see [19]) provides an abstraction of distributed problem solving amongst self-interested agents interacting with each other. In the language of mechanism design, agents are characterised by *types*, which are abstractions of their internal, private beliefs. Given two agents, α, β, the space of possible types for agent $x \in \{\alpha, \beta\}$ is denoted by Θ_x and its type is $\theta_x \in \Theta_x$. Moreover, $\Theta = \Theta_\alpha \times \Theta_\beta$.

Inter-agent interactions have a number of potential *outcomes* $\mathcal{O}$. A given *social choice function* $f : \Theta \rightarrow \mathcal{O}$ characterises what can be deemed to be a *socially* optimal outcome of the interaction. Agents' self-interest is expressed in terms of (private) *utility functions* $v_x : \mathcal{O} \times \Theta_x \rightarrow \mathbb{R}$. A_x is the set of possible actions of agent x and $A = A_\alpha \times A_\beta$. The decision for agent x of which action to perform is given by a *strategy*[4]. Let S_x denote the space of possible strategies for agent x, $S = S_\alpha \times S_\beta$. Then a strategy $s_x \in S_x$ is a function $s_x : \Theta_x \times A \rightarrow A_x$.

Finally, a *mechanism* $\mathcal{M} = (A, g)$ consists of the action space A and an *outcome function* $g : S \rightarrow \mathcal{O}$, where $g(s)$ is the outcome implemented by $\mathcal{M}$ for strategy $s = (s_\alpha, s_\beta)$.

A social choice function specifies the desired goal of an interaction, whereas a mechanism is a means of characterising the agents' behaviour in the interaction. Several characterisations of strategies have been provided as ways to predict how (rational) agents will behave in a mechanism. In particular, a strategy s_x is *dominant* (for x) if it maximises the agent's utility irrespectively of the other agent's strategy. For a mechanism $\mathcal{M} = (A, g)$ and a social choice function f, $\mathcal{M}$ *implements* f iff $g(s) = f(\theta)$, where s is a dominant strategy.

3. GAME SOLUTIONS IN ABA

As a general purpose framework, ABA can be used to represent knowledge in many applications [41]. Here, we give ABA representations of games of the kind presented in Section 1 and show a correspondence between admissible sentences in these frameworks and dominant solutions / Nash equilibria for the games.

Firstly, we specify the languages underlying the ABA frameworks we will use.

DEFINITION 3.1. *Given a game $\langle \Sigma_\alpha, \Sigma_\beta, \pi_\alpha, \pi_\beta \rangle$, the* Dominant/Nash languages *are (respectively):*

- $\mathcal{L}_D = \{d(\sigma_\alpha, \sigma_\beta), nD(\sigma_\alpha, \sigma_\beta), vA(\sigma_\alpha, \sigma_\beta) > vA(\sigma'_\alpha, \sigma'_\beta), vB(\sigma_\alpha, \sigma_\beta) > vB(\sigma'_\alpha, \sigma'_\beta) | \sigma_\alpha, \sigma'_\alpha \in \Sigma_\alpha, \sigma_\beta, \sigma'_\beta \in \Sigma_\beta\}$;
- $\mathcal{L}_N = \{n(\sigma_\alpha, \sigma_\beta), nN(\sigma_\alpha, \sigma_\beta), vA(\sigma_\alpha, \sigma_\beta) > vA(\sigma'_\alpha, \sigma'_\beta), vB(\sigma_\alpha, \sigma_\beta) > vB(\sigma'_\alpha, \sigma'_\beta) | \sigma_\alpha, \sigma'_\alpha \in \Sigma_\alpha, \sigma_\beta, \sigma'_\beta \in \Sigma_\beta\}$.

Intuitively, the sentences in $\mathcal{L}_D$/$\mathcal{L}_N$ can be read as follows: $d(\sigma_\alpha, \sigma_\beta)/n(\sigma_\alpha, \sigma_\beta)$ stands for "$(\sigma_\alpha, \sigma_\beta)$ is a dominant solution/ Nash equilibrium" (respectively); $nD(\sigma_\alpha, \sigma_\beta)/nN(\sigma_\alpha, \sigma_\beta)$ stands for "$(\sigma_\alpha, \sigma_\beta)$ is *not* a dominant solution /Nash equilibrium" (respectively); $vA(\sigma_\alpha, \sigma_\beta) > vA(\sigma'_\alpha, \sigma'_\beta)$ represents that, for agent α, the payoff value of $(\sigma_\alpha, \sigma_\beta)$ is higher than the payoff value of $(\sigma'_\alpha, \sigma'_\beta)$; $vB(\sigma_\alpha, \sigma_\beta) > vB(\sigma'_\alpha, \sigma'_\beta)$ can be read similarly for agent β.

Preferences between strategy pairs (namely sentences of the form $vA(\sigma_\alpha, \sigma_\beta) > vA(\sigma'_\alpha, \sigma'_\beta)$) are heads of *preference rules*, determined by the agents' private payoff functions, as follows:

DEFINITION 3.2. *Given a game $\langle \Sigma_\alpha, \Sigma_\beta, \pi_\alpha, \pi_\beta \rangle$, the* preference rules *$\mathcal{R}^p_\alpha$, for α, and $\mathcal{R}^p_\beta$, for β, are:*

$\mathcal{R}^p_\alpha = \{vA(\sigma_\alpha, \sigma_\beta) > vA(\sigma'_\alpha, \sigma'_\beta) \leftarrow | \sigma_\alpha, \sigma'_\alpha \in \Sigma_\alpha, \sigma_\beta, \sigma'_\beta \in \Sigma_\beta, \pi_\alpha(\sigma_\alpha, \sigma_\beta) > \pi_\alpha(\sigma'_\alpha, \sigma'_\beta)\}$;

$\mathcal{R}^p_\beta = \{vB(\sigma_\alpha, \sigma_\beta) > vB(\sigma'_\alpha, \sigma'_\beta) \leftarrow | \sigma_\alpha, \sigma'_\alpha \in \Sigma_\alpha, \sigma_\beta, \sigma'_\beta \in \Sigma_\beta, \pi_\beta(\sigma_\alpha, \sigma_\beta) > \pi_\beta(\sigma'_\alpha, \sigma'_\beta)\}$.

EXAMPLE 3.1. *(Example 1.1 cont.) For the husband, $\mathcal{R}^p_\alpha$ is:*

$\{vA(m, r) > vA(m, w) \leftarrow,\quad vA(m, r) > vA(f, r) \leftarrow,$
$vA(m, r) > vA(f, w) \leftarrow,\quad vA(f, w) > vA(f, r) \leftarrow,$
$vA(f, w) > vA(m, w) \leftarrow\}$.

For the wife, $\mathcal{R}^p_\beta$ is:

$\{vB(m, r) > vB(m, w) \leftarrow,\quad vB(m, r) > vA(f, r) \leftarrow,$
$vB(f, w) > vB(f, r) \leftarrow,\quad vB(f, w) > vA(m, w) \leftarrow,$
$vB(f, w) > vB(m, r) \leftarrow\}$.

We then define ABA frameworks for capturing dominant solutions.

DEFINITION 3.3. *Given a game $\langle \Sigma_\alpha, \Sigma_\beta, \pi_\alpha, \pi_\beta \rangle$, the* Dominant ABA framework *$\langle \mathcal{L}_D, \mathcal{R}_D, \mathcal{A}_D, \mathcal{C}_D \rangle$ has:*

- *$\mathcal{R}_D$ consisting of $\mathcal{R}^p_\alpha \cup \mathcal{R}^p_\beta$ together with:*[5]
$nD(\sigma_\alpha, \sigma_\beta) \leftarrow vA(\_, \_) > vA(\sigma_\alpha, \sigma_\beta)$,
$nD(\sigma_\alpha, \sigma_\beta) \leftarrow vB(\_, \_) > vB(\sigma_\alpha, \sigma_\beta)$;
- $\mathcal{A}_D = \{d(\sigma_\alpha, \sigma_\beta) | \sigma_\alpha \in \Sigma_\alpha, \sigma_\beta \in \Sigma_\beta\}$;
- *for each $d(\sigma_\alpha, \sigma_\beta) \in \mathcal{A}_D$, $\mathcal{C}_D(d(\sigma_\alpha, \sigma_\beta)) = \{nD(\sigma_\alpha, \sigma_\beta)\}$.*

Note that Dominant ABA frameworks are flat (see Section 2).

The intuition of this definition is the following. Each strategy pair can be *assumed* to be a dominant solution (in that $d(\sigma_\alpha, \sigma_\beta) \in \mathcal{A}_D$ for each $(\sigma_\alpha, \sigma_\beta)$). A pair $(\sigma_\alpha, \sigma_\beta)$ can be deemed to be not dominant iff there is another pair $(\sigma'_\alpha, \sigma'_\beta)$ such that either $\pi_\alpha(\sigma'_\alpha, \sigma'_\beta) > \pi_\alpha(\sigma_\alpha, \sigma_\beta)$ or $\pi_\beta(\sigma'_\alpha, \sigma'_\beta) > \pi_\beta(\sigma_\alpha, \sigma_\beta)$, described by the the definition of contrary and by the two rule schemata in Definition 3.3, respectively (by instantiating $vA(\_, \_)$, $vB(\_, \_)$ to $vA(\sigma'_\alpha, \sigma'_\beta)$, $vB(\sigma'_\alpha, \sigma'_\beta)$, respectively).

Admissible sentences in Dominant ABA frameworks correspond to dominant solutions, as follows.

THEOREM 3.1. *Given a game $\langle \Sigma_\alpha, \Sigma_\beta, \pi_\alpha, \pi_\beta \rangle$, a strategy pair $(\sigma^\star_\alpha, \sigma^\star_\beta)$ is a dominant solution iff $d(\sigma^\star_\alpha, \sigma^\star_\beta)$ is admissible in the Dominant ABA framework.*

[3]ABA dialogues for the grounded and ideal semantics have also been defined in [15], but we only need admissiblity in this work.

[4]In Section 1, including Example 1.1, strategies are considered as actions directly, for ease of presentation.

[5]For brevity, here and in the remainder of the paper, we use *schemata* with variables $(\sigma_\alpha, \sigma_\beta)$ and anonymous variables (_) to represent compactly all rules that can be obtained by instantiating the variables over the appropriate domains.

PROOF. ($\Rightarrow$) To show ($\sigma_\alpha^\star$, $\sigma_\beta^\star$) is a dominant solution is to show for all other strategy pairs $(\sigma_\alpha, \sigma_\beta)$, $\pi_\alpha(\sigma_\alpha^\star, \sigma_\beta^\star) \geq \pi_\alpha(\sigma_\alpha, \sigma_\beta)$ and $\pi_\beta(\sigma_\alpha^\star, \sigma_\beta^\star) \geq \pi_\beta(\sigma_\alpha, \sigma_\beta)$. Suppose otherwise, namely there is $(\sigma'_\alpha, \sigma'_\beta)$ such that $\pi_\alpha(\sigma'_\alpha, \sigma'_\beta) > \pi_\alpha(\sigma_\alpha^\star, \sigma_\beta^\star)$. Thus, by Definition 3.2, $vA(\sigma'_\alpha, \sigma'_\beta) > vA(\sigma_\alpha^\star, \sigma_\beta^\star) \leftarrow \in \mathcal{R}_\alpha^p$. Given the rule $nD(\sigma_\alpha, \sigma_\beta) \leftarrow vA(\sigma'_\alpha, \sigma'_\beta) > vA(\sigma_\alpha, \sigma_\beta)$ in the Dominant ABA framework, let $\sigma_\alpha = \sigma_\alpha^\star$ and $\sigma_\beta = \sigma_\beta^\star$; then argument $\{\} \vdash nD(\sigma_\alpha^\star, \sigma_\beta^\star)$ attacks $\{d(\sigma_\alpha^\star, \sigma_\beta^\star)\} \vdash d(\sigma_\alpha^\star, \sigma_\beta^\star) = A$. Therefore A is not admissible, so $d(\sigma_\alpha^\star, \sigma_\beta^\star)$ is not admissible either. Contradiction.
($\Leftarrow$) To show $d(\sigma_\alpha^\star, \sigma_\beta^\star)$ is admissible is to show that there is a set of argument Δ such that $\{d(\sigma_\alpha^\star, \sigma_\beta^\star)\} \vdash d(\sigma_\alpha^\star, \sigma_\beta^\star) = A \in \Delta$ and Δ is conflict-free and counterattacks all attacking arguments. Since $\mathcal{C}(d(\sigma_\alpha^\star, \sigma_\beta^\star)) = \{nD(\sigma_\alpha^\star, \sigma_\beta^\star)\}$, arguments attacking A are of the form $\{\} \vdash nD(\sigma_\alpha^\star, \sigma_\beta^\star)$. Since $(\sigma_\alpha^\star, \sigma_\beta^\star)$ is a dominant solution, there is no $(\sigma'_\alpha, \sigma'_\beta)$ such that $\pi_\alpha(\sigma'_\alpha, \sigma'_\beta) > \pi_\alpha(\sigma_\alpha^\star, \sigma_\beta^\star)$. Thus $vA(\sigma'_\alpha, \sigma'_\beta) > vA(\sigma_\alpha^\star, \sigma_\beta^\star) \leftarrow \notin \mathcal{R}_\alpha^p$. Therefore $\{\} \vdash nD(\sigma_\alpha^\star, \sigma_\beta^\star)$ is not an argument and $\Delta = \{A\}$ is not attacked. Thus $d(\sigma_\alpha^\star, \sigma_\beta^\star)$ is admissible. $\square$

EXAMPLE 3.2. *(Example 3.1 cont.) There is no dominant solution in this game. Indeed, for any $(\sigma_\alpha^\star, \sigma_\beta^\star)$, $d(\sigma_\alpha^\star, \sigma_\beta^\star)$ is not admissible in the Dominant ABA framework. For instance, for $\sigma_\alpha^\star = m$ and $\sigma_\beta^\star = r$, the argument $\{d(m, r)\} \vdash d(m, r)$ is attacked by the argument $\{\} \vdash nD(m, r)$ (this argument uses the rule $vB(f, w) > vB(m, r) \leftarrow$, and cannot be counter-attacked). Thus, an argumentative reading of the process for sanctioning $d(\sigma_\alpha^\star, \sigma_\beta^\star)$ as non-dominant is:*

1. *$\{d(m, r)\} \vdash d(m, r)$: let us assume that (m, r) is dominant;*
2. *$\{\} \vdash nD(m, r)$ (using $vB(f, w) > vB(m, r) \leftarrow$) attacks the assumption/argument supporting the assumption:[6] (m, r) is not dominant because (f, w) has a higher payoff value than (m, r) for agent β.*

We can also easily define ABA frameworks for capturing Nash Equilibria as follows.

DEFINITION 3.4. *Given a game $\langle \Sigma_\alpha, \Sigma_\beta, \pi_\alpha, \pi_\beta \rangle$, the* Nash ABA framework *$\langle \mathcal{L}_N, \mathcal{R}_N, \mathcal{A}_N, \mathcal{C}_N \rangle$ has:*

- *$\mathcal{R}_N$ consisting of $\mathcal{R}_\alpha^p \cup \mathcal{R}_\beta^p$ together with:*
  $nN(\sigma_\alpha, \sigma_\beta) \leftarrow vA(\_, \sigma_\beta) > vA(\sigma_\alpha, \sigma_\beta)$,
  $nN(\sigma_\alpha, \sigma_\beta) \leftarrow vB(\sigma_\alpha, \_) > vB(\sigma_\alpha, \sigma_\beta)$;
- *$\mathcal{A}_N = \{n(\sigma_\alpha, \sigma_\beta) | \sigma_\alpha \in \Sigma_\alpha, \sigma_\beta \in \Sigma_\beta\}$;*
- *for each $n(\sigma_\alpha, \sigma_\beta) \in \mathcal{A}_N$, $\mathcal{C}_N(n(\sigma_\alpha, \sigma_\beta)) = \{nN(\sigma_\alpha, \sigma_\beta)\}$.*

Note that Nash ABA frameworks are also flat (see Section 2).

The intuition behind this definition is analogous to that behind Definition 3.3, except that upon checking Nash equilibria, one agent's payoff in a strategy pair only needs to be compared to strategy pairs with the other agent playing the same strategy (see the two rule schemata in Definition 3.4).

THEOREM 3.2. *Given a game $\langle \Sigma_\alpha, \Sigma_\beta, \pi_\alpha, \pi_\beta \rangle$, a strategy pair ($\sigma_\alpha$, σ_β) is a Nash equilibrium iff $n(\sigma_\alpha, \sigma_\beta)$ is admissible in the Nash ABA framework.*

[6]Note that, in flat ABA frameworks, arguments supporting assumptions necessarily use no rules and are supported by the assumptions themselves.

The proof of Theorem 3.2 is similar to the one of Theorem 3.1.

EXAMPLE 3.3. *(Example 3.1 cont.) The two Nash Equilibria are: (m, r) and (f, w); indeed, both $n(m, r)$ and $n(f, w)$ are admissible in the Nash ABA framework. For instance, again for $\sigma_\alpha^\star = m$ and $\sigma_\beta^\star = r$, we can see that the argument $\{n(m, r)\} \vdash n(m, r)$ is not attacked by any argument as there is no rule with head $vA(f, r) > vA(m, r)$ or $vB(m, w) > vB(m, r)$. Argumentatively, (m, r) is a Nash Equilibrium since α does not think (f, r) has a higher payoff value than (m, r) and β does not think (m, w) has a higher payoff value than (m, r).*

Note that as we can see that both Dominant ABA frameworks and Nash ABA frameworks are identical except for some rules, we could use a unified framework by taking the "union" of the two. We leave them seperate in this work for clarity.

4. GAME SOLUTIONS VIA ABA DIALOGUES

In addition to computing solutions with complete information statically, dominant solutions and Nash equilibria can be computed dynamically, with only partial information held by each agent, through ABA dialogues. More specifically, each agent is only aware of its own *half* of the overall payoff matrix and the corresponding preference rules; however, both agents hold the rules for computing solutions (represented in the shared schemata, assumptions and contraries in Definitions 3.3 and 3.4, respectively).

In the remainder of the paper, unless specified otherwise, we assume as given a game $\langle \Sigma_\alpha, \Sigma_\beta, \pi_\alpha, \pi_\beta \rangle$ and corresponding Dominant / Nash ABA framework $\langle \mathcal{L}_X, \mathcal{R}_X, \mathcal{A}_X, \mathcal{C}_X \rangle$ (for $X = D$ / N respectively). Then, the ABA frameworks held by α and β will be denoted by $\alpha = \langle \mathcal{L}_X, \mathcal{R}_\alpha, \mathcal{A}_\alpha, \mathcal{C}_\alpha \rangle$ and $\beta = \langle \mathcal{L}_X, \mathcal{R}_\beta, \mathcal{A}_\beta, \mathcal{C}_\beta \rangle$, such that:
- $\mathcal{R}_\alpha = \mathcal{R}_X \setminus \mathcal{R}_\beta^p$; • $\mathcal{R}_\beta = \mathcal{R}_X \setminus \mathcal{R}_\alpha^p$;
- $\mathcal{A}_\alpha = \mathcal{A}_\beta = \mathcal{A}_X$; • $\mathcal{C}_\alpha(a) = \mathcal{C}_\beta(a) = \mathcal{C}_X(a)$, for all $a \in \mathcal{A}_X$.

Thus, the ABA framework held by an agent contains exactly the same information as the available strategies and (private) payoff function, as well as information as to which notion of solution the agents are striving towards.

Through ABA dialogues, while exchanging relevant information, agents can identify solutions even if they hold only incomplete information. *Dominant / Nash dialogues* for computing dominant solutions and Nash Equilibria use specific legal-move functions:

DEFINITION 4.1. *The* Dominant Legal-move *function λ_d is such that given a dialogue δ, each utterance $u \in \lambda_d(\delta)$ has content in one of the following forms (for some $\sigma_\alpha, \sigma'_\alpha \in \Sigma_\alpha, \sigma_\beta, \sigma'_\beta \in \Sigma_\beta$):*
- $claim(d(\sigma_\alpha, \sigma_\beta))$,
- $asm(d(\sigma_\alpha, \sigma_\beta))$, • $ctr(d(\sigma_\alpha, \sigma_\beta), nD(\sigma_\alpha, \sigma_\beta))$,
- $rl(nD(\sigma_\alpha, \sigma_\beta) \leftarrow vA(\sigma'_\alpha, \sigma'_\beta) > vA(\sigma_\alpha, \sigma_\beta))$,
- $rl(nD(\sigma_\alpha, \sigma_\beta) \leftarrow vB(\sigma'_\alpha, \sigma'_\beta) > vB(\sigma_\alpha, \sigma_\beta))$,
- $rl(vA(\sigma'_\alpha, \sigma'_\beta) > vA(\sigma_\alpha, \sigma_\beta) \leftarrow)$,
- $rl(vB(\sigma'_\alpha, \sigma'_\beta) > vB(\sigma_\alpha, \sigma_\beta) \leftarrow)$.

The Nash Legal-move *function λ_n is defined as λ_d with*
- *$d(\sigma_\alpha, \sigma_\beta)$ replaced by $n(\sigma_\alpha, \sigma_\beta)$;*
- *$nD(\sigma_\alpha, \sigma_\beta)$ replaced by $nN(\sigma_\alpha, \sigma_\beta)$;*
- *the two rule schemata with heads $nD(\sigma_\alpha, \sigma_\beta)$ replaced by their counterparts in the Nash ABA framework with heads $nN(\sigma_\alpha, \sigma_\beta)$.*

A dialogue δ is a Dominant / Nash dialogue *iff δ is compatible with λ_d / λ_n, respectively. The strategy pair in the claim of δ is called the* claim pair.

Definition 4.1 gives the dialectical counterpart of Definitions 3.3 and 3.4. It specifies the allowed utterances for agents in Dominant /

Nash dialogues. However, it does not force agents to disclose their private preferences. In particular, the last two rules in the definition of λ_d may not be preference rules in the ABA framework of the utterer (in other words, they may be lies). As seen in Section 2, the thorough strategy-move function ϕ_h can be used to dictate agents to be *honest*, by disclosing all relevant information they hold.

THEOREM 4.1. *For any Dominant / Nash dialogue δ with claim pair $(\sigma^\star_\alpha, \sigma^\star_\beta)$ constructed with ϕ_h, δ is successful iff $(\sigma^\star_\alpha, \sigma^\star_\beta)$ is a dominant solution / Nash equilibrium.*

PROOF. Follows from Theorems 3.1 and 3.2, and Theorem 1 in [13]. $\square$

Note that Theorem 4.1 sanctions both directions of the "equivalence" between dialogues and games.

EXAMPLE 4.1. *(Example 3.3 cont.) We illustrate the Nash dialogue $\mathcal{D}^\alpha_\beta(n(m, r))$ being successful as follows:*

$\langle \alpha, \beta, 0, claim(n(m, r)), 1\rangle$
$\langle \alpha, \beta, 1, asm(n(m, r)), 2\rangle$
$\langle \alpha, \beta, 2, ctr(n(m, r), nN(m, r)), 3\rangle$
$\langle \beta, \alpha, 3, rl(nN(m, r) \leftarrow vA(f, r) > vA(m, r)), 4\rangle$
$\langle \alpha, \beta, 0, \pi, 5\rangle$
$\langle \alpha, \beta, 3, rl(nN(m, r) \leftarrow vB(m, w) > vB(m, r)), 6\rangle$
$\langle \beta, \alpha, 0, \pi, 7\rangle$

This is a Nash dialogue as it is compatible with the Nash legal-move function given in Definition 4.1. A natural language reading of this dialogue is given in Section 1. In utterances 1-3, α assumes (m, r) being a Nash equilibrium with contrary $nN(m, r)$. β utters the two rules for $nN(m, r)$ in utterances 4, 5. Since α does not prefer (f, r) to (m, r) nor β prefers (m, w) to (m, r), no more utterances are made. Thus, jointly the agents verify that (m, r) is a Nash equilibrium. Overall, the only argument they generate and exchange (implicitly) is $\{n(m, r)\} \vdash n(m, r)$ (this is the only argument in the ABA framework drawn from the dialogue). Since there is no attack against this argument (in the ABA framework drawn from the dialogue), the claim is admissible and the dialogue is successful.

Example 4.1 also illustrates a difference between using structured rather than abstract argumentation (AA) as the argumentation framework underlying interactions between agents (in general and in the context of determining game solutions). From an AA view point, the dialogue in Example 4.1 solely introduces an argument (for (m, r) being a Nash equilibrium). The dialogue also attempts to generate attacks against this argument (one attempt is given by utterances 3 and 4, another by utterances 3 and 6). The attempts allow to share additional information (as to the absence of preferences in this case). This may be useful, for example, when agents have only partial information about the world before engaging in a dialogue, and may benefit from information from the other agents to even determine their payoff functions. We leave this extension to our current, standard game model as future work.

The following example further illustrates how ABA dialogues support the sharing of information, but in the case of non-successful Dominant dialogues.

EXAMPLE 4.2. *(Example 3.2 cont.) The following Dominant dialogue $\mathcal{D}^\alpha_\beta(d(m, r))$, attempting to determine whether $d(m, r)$ is a dominant solution, is not successful:*

$\langle \alpha, \beta, 0, claim(d(m, r)), 1\rangle$
$\langle \alpha, \beta, 1, asm(d(m, r)), 2\rangle$
$\langle \alpha, \beta, 2, ctr(d(m, r), nD(m, r)), 3\rangle$
$\langle \beta, \alpha, 3, rl(nD(m, r) \leftarrow vA(m, w) > vA(m, r)), 4\rangle$
$\langle \alpha, \beta, 3, rl(nD(m, r) \leftarrow vA(f, r) > vA(m, r)), 5\rangle$
$\langle \alpha, \beta, 3, rl(nD(m, r) \leftarrow vA(f, w) > vA(m, r)), 6\rangle$
$\langle \alpha, \beta, 3, rl(nD(m, r) \leftarrow vB(m, w) > vB(m, r)), 7\rangle$
$\langle \alpha, \beta, 3, rl(nD(m, r) \leftarrow vB(f, r) > vB(m, r)), 8\rangle$
$\langle \alpha, \beta, 3, rl(nD(m, r) \leftarrow vB(f, w) > vB(m, r)), 9\rangle$
$\langle \beta, \alpha, 9, rl(vB(f, w) > vB(m, r) \leftarrow), 10\rangle$
$\langle \alpha, \beta, 0, \pi, 11\rangle$
$\langle \beta, \alpha, 0, \pi, 12\rangle$

This is a Dominant dialogue as it is compatible with the Dominant legal-move function of Definition 4.1. It constructs two arguments: $\{d(m, r)\} \vdash d(m, r)$ and $\{\} \vdash nD(m, r)$; the latter attacks the former. Therefore, the claim $d(m, r)$ is not admissible (in the ABA framework drawn from the dialogue) and the dialogue is not successful. Nonetheless, after the dialogue the agents share information of the preference by β of (f, w) over (m, r) (although not the utility of either for β). Whether this is an actual preference of β depends on whether β has been truthfully honest in the dialogue. This is the topic of the next section.

5. DIALECTICAL STRATEGIES FOR ABA DIALOGUES

The dialogue in Example 4.1 is constructed with ϕ_h (see Section 2). However, if we consider agents being self interested, we need to ask:

> *What are the conditions for agents to be honest in dialogues?*

We pose this question in a game theoretic setting and analyse agent behaviours in dialogues understood as games themselves. Thus, we introduce a *meta-level dialectical game* on top of the original *underlying game* (e.g. the Dinner party game in Example 1.1). Note that although both agents carry information from the underlying game into the meta-level game, they do not necessarily carry the payoffs. Indeed, *designing* appropriate utilities for the meta-level game is needed to ensure good behaviour (truthfulness and disclosure) from agents.

We formulate (reverse) game-theoretic notions for dialogues as follows. We first define the types of agents as their strategies and payoffs in the underlying game, represented within their (Dominant or Nash) ABA frameworks.

DEFINITION 5.1. *The* types *of agents α and β are $\theta_\alpha = \alpha$ and $\theta_\beta = \beta$, respectively.*

In ABA dialogues, agents put forward claims, rules, assumptions and contraries. We hence view sequences of utterances as agents' actions in the meta-level dialectical game and define strategies in the meta-level dialectical game as functions determining utterances in dialogues (that can be equated with strategy-move functions, as we shall see). Formally:

DEFINITION 5.2. *The* action space *for agent $x \in \{\alpha, \beta\}$ is* $2^{\mathcal{U}^x}$.

DEFINITION 5.3. *A* strategy for agent $x \in \{\alpha, \beta\}$ *(in the meta-level dialectical game) is a function s^δ_x, for some dialogue δ, such that $s^\delta_x(\theta_x, \delta) = \{u | u = \langle x, \_, \_, \_, \_\rangle \in \delta\}$.*

Since s^δ_x returns the set of utterances made by x in δ, which is determined by the strategy-move function ϕ used by x, we can equate strategies in the meta-level dialectical game with strategy-move functions. If δ is constructed with ϕ_h, then s^δ_x is called the *honest strategy* (in the meta-level dialectical game).

Since a dialogue can either be successful or not, we let the meta-level outcome depend on the dialogue outcome such that if the dialogue is successful, then the outcome is the claim pair; otherwise, the outcome is the empty set, formally:

DEFINITION 5.4. *Given a dialogue $\delta = \mathcal{D}^x_y(X(\sigma^\star_\alpha, \sigma^\star_\beta)), x, y \in \{\alpha, \beta\}, x \neq y, X \in \{d, n\}$, the* outcome function $g(\delta)$ *is such that:*

$$g(\delta) = \begin{cases} (\sigma^\star_\alpha, \sigma^\star_\beta) & \textit{if } \delta \textit{ is successful,} \\ \{\} & \textit{otherwise.} \end{cases}$$

Note that success of a dialogue amounts to admissibility of its claim in the ABA framework *drawn from the dialogue* (see Section 2). Thus, if the agents are not truthful within the dialogue, the claim pair may not be a dominant solution/Nash equilibrium.

To define the utility function, we first define the *honesty reward*, a means to encourage agents' truthfulness. Inspired by Vickrey auctions[23], in Dominant dialogues, the honesty reward for each agent is set to the second highest payoff of the underlying game; in Nash dialogues, the honesty reward is the second highest payoff *given the other agent's behaviour* in the underlying game. Formally:

DEFINITION 5.5. *The* honesty reward $\mathcal{W}_x \in \mathbb{R}$ *(for $x \in \{\alpha, \beta\}$) is given as follows:*

- *For Dominant dialogues, let $\langle v^1_x, v^2_x, \ldots \rangle$ be the list of of payoffs of agent x such that $v^1_x > v^2_x > \ldots$; then $\mathcal{W}_x = v^2_x$;*
- *For Nash dialogues, let $(\sigma^\star_\alpha, \sigma^\star_\beta)$ be the claim pair, and*
  - *let $\langle v^1_\alpha, v^2_\alpha, \ldots \rangle$ be the list of payoffs of α when β plays $\sigma^\star_\beta$ in the underlying game, such that $v^1_\alpha > v^2_\alpha > \ldots$; then $\mathcal{W}_\alpha = v^2_\alpha$;*
  - *let $\langle v^1_\beta, v^2_\beta, \ldots \rangle$ be the list of payoffs of β when α plays $\sigma^\star_\alpha$ in the underlying game, such that $v^1_\beta > v^2_\beta > \ldots$; then $\mathcal{W}_\beta = v^2_\beta$.*

An agent's utility is defined so that the agent either gets its payoff in the underlying game (when the dialogue is successful) or its honesty reward (otherwise):

DEFINITION 5.6. *The* utility of agent $x \in \{\alpha, \beta\}$ in dialogue δ *is v_x such that:*

$$v_x = \begin{cases} \pi_x(\sigma^\star_\alpha, \sigma^\star_\beta) & \textit{if } g(\delta) = (\sigma^\star_\alpha, \sigma^\star_\beta), \\ \mathcal{W}_x & \textit{otherwise.} \end{cases}$$

Generally speaking, the honesty reward is set to ensure that being truthful yields a higher utility than being dishonest for both agents while participating in Dominant / Nash dialogues. Indeed, with utilities defined as such, we show that truthfully disclosing all relevant information is the best option for both agents. Formally:

THEOREM 5.1. *Given a Dominant / Nash dialogue δ, the honest strategy s^δ_x is dominant (for agent x).*

PROOF. (Sketch.) We show the case for Dominant dialogues- The other case, Nash dialogues, is similar. To show that s^δ_x is a dominant strategy is to show that being thorough gives the highest utility. There are four possible cases in total. (Below, we say that a strategy pair is *dominant* for x if it gives the highest payoff to x.[7])

C1: $(\sigma^\star_\alpha,\sigma^\star_\beta)$ is dominant for both agents;
C2: $(\sigma^\star_\alpha,\sigma^\star_\beta)$ is dominant for β, but not α;
C3: $(\sigma^\star_\alpha,\sigma^\star_\beta)$ is dominant for α, but not β;
C4: $(\sigma^\star_\alpha,\sigma^\star_\beta)$ is not dominant for either α or β.

[7]Note that *dominant* is overloaded in this paper. A pair of strategies can be a dominant solution as defined in Section 1; a single strategy can be dominant as defined at the end of Section 2; and a pair of strategies can be dominant for a single agent as used here.

C1	β		
		H	**D**
α	**H**	$(\sigma^\star_\alpha,\sigma^\star_\beta)$	{}
	D	{}	{}

C2	β		
		H	**D**
α	**H**	{}	{}
	D	$(\sigma^\star_\alpha,\sigma^\star_\beta)$	{}

C3	β		
		H	**D**
α	**H**	{}	$(\sigma^\star_\alpha,\sigma^\star_\beta)$
	D	{}	{}

C4	β		
		H	**D**
α	**H**	{}	{}
	D	{}	$(\sigma^\star_\alpha,\sigma^\star_\beta)$

Table 2: Four possible Cases C1 - C4.

Agents can either be *honest* (**H**) or *dishonest* (**D**). Dialogues can either return its claim pair (if it is successful) or the empty set (otherwise). The dialectical interaction can be viewed as games summarised in Table 2.

C1: In this case, $(\sigma^\star_\alpha, \sigma^\star_\beta)$ is a dominant solution, honesty from both agents ensures a successful dialogue (Theorem 4.1), giving the top left cell (**H, H**). Thus, by Definition 5.6, honesty from both agents gives the highest possible utility, $\pi_x(\sigma^\star_\alpha, \sigma^\star_\beta)$ for $x \in \{\alpha, \beta\}$. Moreover, regardless what the other agent does, being honest yields no less utility.

C2: We first explain how this table is computed.

Since $(\sigma^\star_\alpha, \sigma^\star_\beta)$ is not dominant for α, there is some strategy pair $(\sigma_\alpha, \sigma_\beta)$ such that $\pi_\alpha(\sigma_\alpha, \sigma_\beta) > \pi_\alpha(\sigma^\star_\alpha, \sigma^\star_\beta)$, which gives $vA(\sigma_\alpha, \sigma_\beta) > vA(\sigma^\star_\alpha, \sigma^\star_\beta) \leftarrow$ in $\mathcal{R}^p_\alpha$ (Definition 3.2). Thus, as long as α is honest, by uttering this rule in δ, δ will be unsuccessful and return $\{\}$ (the top row in the C2 table).

If α is dishonest, i.e. α hides this rule in δ, while β is being honest (resulting in the cell (**H, D**)), since $(\sigma^\star_\alpha, \sigma^\star_\beta)$ is dominant for β, β has no rule to attack $d(\sigma^\star_\alpha, \sigma^\star_\beta)$, thus δ will be successful and the outcome of δ is the claim pair $(\sigma^\star_\alpha, \sigma^\star_\beta)$ (lower left cell in the C2 table).

If both α and β are dishonest (**D, D**), β lies about having $(\sigma_\alpha, \sigma_\beta)$ such that $vB(\sigma_\alpha, \sigma_\beta) > vB(\sigma^\star_\alpha, \sigma^\star_\beta) \leftarrow$ is uttered in δ and δ is unsuccessful and returns $\{\}$ (lower right cell in the C2 table).

With the C2 table constructed as such, we observe that α should choose to be honest (i.e. choose to be in the top row) to receive $\mathcal{W}_\alpha$ as, by Definition 5.5, $\mathcal{W}_\alpha \geq \pi_\alpha(\sigma^\star_\alpha, \sigma^\star_\beta)$ (since $(\sigma^\star_\alpha, \sigma^\star_\beta)$ is not dominant for α). β should also choose to be honest as β either receives $\mathcal{W}_\beta$ (when α is honest, resulting in (**H, H**)) or $\pi_\beta(\sigma^\star_\alpha, \sigma^\star_\beta)$ (when α lies, resulting in (**D, H**)). Since $(\sigma^\star_\alpha, \sigma^\star_\beta)$ is dominant for β, $\pi_\beta(\sigma^\star_\alpha, \sigma^\star_\beta) > \mathcal{W}_\beta$. Thus, being honest is also dominant for β.

C3: This case is case **C2** with the two agents swapped.

C4: Honesty gives both agents $\mathcal{W}_x$, otherwise each receives either $\pi_x(\sigma^\star_\alpha, \sigma^\star_\beta)$ or $\mathcal{W}_x$, $x \in \{\alpha, \beta\}$. Since $(\sigma^\star_\alpha, \sigma^\star_\beta)$ is not dominant for the agents, by Definition 5.5, $\mathcal{W}_x \geq \pi_x(\sigma^\star_\alpha, \sigma^\star_\beta)$.

In all cases, the honest strategy dominates. □

We observe that the honesty reward as defined in Definition 5.5 gives enough incentive to both agents to be honest so hiding information or inserting false information does not bring higher utility to either of the two. Moreover, from the proof of Theorem 5.1, we observe that the setting for the honesty reward is very crucial to this result. Indeed, the following lemma holds.

LEMMA 5.1. *Given a Dominant / Nash dialogue δ, the honest strategy s^δ_x is not dominant (for agent x) if the honesty reward $\mathcal{W}_x$ is such that $\mathcal{W}_x < v^2_x$, for v^2_x as in Definition 5.5.*

PROOF. (Sketch.) Again, we only show the Dominant case as the Nash case is similiar. As in the proof of Theorem 5.1, we need to consider four possible cases C1-C4. The outcomes remain the same as in Table 2. In the case of C2, suppose that $\pi_\alpha(\sigma^\star_\alpha, \sigma^\star_\beta) =$

v_α^2. If $\mathcal{W}_x < v_x^2$, then α is better off being dishonest and hiding the rule $vA(\sigma_\alpha, \sigma_\beta) > vA(\sigma_\alpha^\star, \sigma_\beta^\star) \leftarrow$ to make δ successful. By doing so, α could receive v_α^2 (if β is honest), which is better than $\mathcal{W}_\alpha$, which is the utility of α while being honest (δ would not be successful when α is honest). Thus, the honest strategy s_x^δ is not dominant. □

6. DIALOGUE SEQUENCES FOR FINDING SOLUTIONS

A single Dominant / Nash dialogue only tests if a single strategy pair is a dominant solution or a Nash equilibrium. To identify all dominant solutions / Nash equilibria, multiple dialogues are needed. We define *dialogue sequences* as follows.

DEFINITION 6.1. *A* (dialogue) sequence *S between α, β is a set of dialogues $S = \{\mathcal{D}_{x_1}^{y_1}(\chi_1), \ldots, \mathcal{D}_{x_n}^{y_n}(\chi_n)\}$ such that $n > 0$ and for all $i, j = 1, \ldots, n$, $x_i, y_i \in \{\alpha, \beta\}, x_i \neq y_i$ and if $i \neq j$ then $\chi_i \neq \chi_j$.*

A sequence is a dominant / Nash *sequence iff (1) all dialogues in the sequence are Dominant / Nash dialogues (respectively); (2) for all $\sigma_\alpha \in \Sigma_\alpha$ and $\sigma_\beta \in \Sigma_\beta$, ($\sigma_\alpha$, σ_β) is the claim pair for some dialogue in the sequence; and (3) all claims of dialogues in the sequences are claim pairs (σ_α, σ_β), for $\sigma_\alpha \in \Sigma_\alpha$ and $\sigma_\beta \in \Sigma_\beta$.*

Here, we only consider Dominant or Nash sequences.

We do not repeat the definition for action space but assume the one given in Definition 5.2 remains unchanged for sequences.

Similarly to Definition 5.3, we define *sequence strategies* as returning utterances in (all dialogues in) the sequence. Note that this setting also effectively equates sequence strategies with strategy-move functions used in dialogues in the sequence.

DEFINITION 6.2. *For any sequence $S = \{\delta_1, \ldots, \delta_n\}$ between α and β, the* sequence strategy *s_x^S for $x \in \{\alpha, \beta\}$ is such that $s_x^S(\theta_x, S) = \{u | u = \langle x, \_, \_, \_, \_\rangle \in \delta_i, 1 \leq i \leq n\}$.*

We refer to a sequence strategy s_x^S as *honest* iff the strategy-move function used by x in its dialogues is ϕ_h.

The outcome function g wrt a sequence is as follows.

DEFINITION 6.3. *Given a sequence S, the* outcome function *g_X, for $X = D/N$, is such that $g_X(S) = \{(\sigma_\alpha^\star, \sigma_\beta^\star) | Y(\sigma_\alpha^\star, \sigma_\beta^\star)$ is the topic of $\delta \in S$, for $Y = d/n$, respectively, and δ is successful$\}$.*

We continue to use the utility setting for single dialogues given in Definition 5.6 and let the utility of a sequence be the sum of the utilities of dialogues in the sequence, defined as follows.

DEFINITION 6.4. *The* utility *of agent $x \in \{\alpha, \beta\}$ in sequence $S = \{\delta_1, \ldots, \delta_n\}$ is $sv_x = \sum_{i=1}^{n} v_x^i$, where, for $i = 1 \ldots n$, v_x^i is the utility of x in δ_i, as in Definition 5.6.*

As a corollary to Theorem 5.1, the following holds.

COROLLARY 6.1. *Given a Dominant / Nash sequence S, the honest sequence strategy s_x^S is dominant (for agent x).*

We define the Dominant / Nash social choices as follows.

DEFINITION 6.5. *The* Dominant social choice *function is defined as*

$f_d(\theta_\alpha, \theta_\beta) = \{(\sigma_\alpha^\star, \sigma_\beta^\star) | (\sigma_\alpha^\star, \sigma_\beta^\star)$ *is a dominant solution, for* $\sigma_\alpha^\star \in \Sigma_\alpha$, $\sigma_\beta^\star \in \Sigma_\beta$ $\}$.

The Nash social choice *function is defined as*

$f_n(\theta_\alpha, \theta_\beta) = \{(\sigma_\alpha^\star, \sigma_\beta^\star) | (\sigma_\alpha^\star, \sigma_\beta^\star)$ *is a Nash equilibrium, for* $\sigma_\alpha^\star \in \Sigma_\alpha$, $\sigma_\beta^\star \in \Sigma_\beta$ $\}$.

Namely, the Dominant / Nash social choice function returns all dominant solutions / Nash equilibria (respectively) in the underlying game. The following theorem holds.

THEOREM 6.1. *The mechanism $\mathcal{M} = (A, g_X)$, where A is the action space, as given in Definition 5.2, and g_X is the outcome function, as given in Definition 6.3, with $X = D/N$,* implements *the social choice function f_d / f_n (respectively).*

PROOF. (Sketch.) This is to show that both agents reveal their true types in dialogue sequences. Since Corollary 6.1 shows that being honest is the dominant strategy, this holds. □

The following example illustrates the use of Nash sequences as a mechanism implementing the Nash social choice function.

EXAMPLE 6.1. *(Example 4.1 cont.) To find all Nash equilibria in the game of Example 1.1, four Nash dialogues, forming a Nash sequence, are conducted with claims $n(m, r)$, $n(m, w)$, $n(f, r)$ and $n(f, w)$, respectively (with both agents use the honest sequence strategy). Example 4.1 gives one of the four dialogues. Dialogues $\mathcal{D}_\beta^\alpha(n(m, w))$ and $\mathcal{D}_\beta^\alpha(n(f, w))$ are as follows:*

$\langle \alpha, \beta, 0, claim(n(m, w)), 1\rangle$
$\langle \alpha, \beta, 1, asm(n(m, w)), 2\rangle$
$\langle \alpha, \beta, 2, ctr(n(m, w), nN(m, w)), 3\rangle$
$\langle \beta, \alpha, 3, rl(nN(m, w) \leftarrow vA(f, w) > vA(f, w)), 4\rangle$
$\langle \alpha, \beta, 4, rl(vA(f, w) > vA(m, w) \leftarrow), 5\rangle$
$\langle \alpha, \beta, 3, rl(nN(m, w) \leftarrow vB(m, r) > vB(m, w)), 6\rangle$
$\langle \beta, \alpha, 6, rl(vB(m, r) > vB(m, w) \leftarrow), 7\rangle$
$\langle \beta, \alpha, 0, \pi, 8\rangle$
$\langle \alpha, \beta, 0, \pi, 9\rangle$

$\langle \alpha, \beta, 0, claim(n(f, w)), 1\rangle$
$\langle \alpha, \beta, 1, asm(n(f, w)), 2\rangle$
$\langle \alpha, \beta, 2, ctr(n(f, w), nN(f, w)), 3\rangle$
$\langle \beta, \alpha, 3, rl(nN(f, w) \leftarrow vA(m, w) > vA(f, w)), 4\rangle$
$\langle \alpha, \beta, 0, \pi, 5\rangle$
$\langle \alpha, \beta, 3, rl(nN(f, w) \leftarrow vB(f, r) > vB(f, w)), 6\rangle$
$\langle \beta, \alpha, 0, \pi, 7\rangle$

Clearly, $\mathcal{D}_\beta^\alpha(n(m, w))$ is not successful, whereas $\mathcal{D}_\beta^\alpha(n(f, w))$ is. It is also easy to see that $\mathcal{D}_\beta^\alpha(n(f, r))$ cannot be successful. Thus, the outcome of this sequence is $\{(m, r), (f, w)\}$, which is the same as the output of the Nash social choice function.

The notions we have introduced in Sections 5 and 6 are summarised in Table 3.

7. RELATED WORKS

Dung's seminal work [9] introduces AA frameworks and studies, as an illustration of the use of AA, modelling n-person games and stable marriage problems with AA. The first half of this paper (sections 3 and 4) continues this spirit, yet instead of modelling a single type of game, we give an ABA modelling for any game in normal form.

[36] and [29] have introduced Argumentation Mechanism Design. Their works give conditions on argumentation framework structures that ensure truthfulness of agents. Instead of putting restrictions on argumentation frameworks, we look at utility settings for enforcing truthfulness. Furthermore, [36] and [29] used the AA framework whereas our work is based on ABA and ABA dialogues.

[35] present examples of logical mechanism design. The main point of their work is to demonstrate introducing mechanism design as a tool in the design of logical inference procedures, whereas our paper focuses on directly applying mechanism design in ABA dialogues.

Mechanism Design Concepts	$(x \in \{\alpha, \beta\})$	Argumentation Dialogues	Dialogue Sequences
Type space (of agent x)	Θ_x	$\Sigma_x, \pi_x, \langle \mathcal{L}, \mathcal{R}_x, \mathcal{A}_x, \mathcal{C}_x \rangle$ (Defn 5.1)	*as in dialogues*
Outcomes	$\mathcal{O}$	$(\sigma_\alpha^\star, \sigma_\beta^\star)$ or $\{\}$ (Defn 5.4)	$\{(\sigma_\alpha^1, \sigma_\beta^1), (\sigma_\alpha^2, \sigma_\beta^2), \ldots\}$ (Defn 6.3)
Utility function (for agent x)	$v_x : \mathcal{O} \times \theta_x \to \mathbb{R}$	$\pi_x(\sigma_\alpha^\star, \sigma_\beta^\star)$ or $\mathcal{W}_x$ (Defn 5.6)	$sv_x = \sum v_x^i$ (Defn 6.4)
Action space	$A = A_\alpha \times A_\beta$	$2^{\mathcal{U}^\alpha} \times 2^{\mathcal{U}^\beta}$ (Defn 5.2)	*as in dialogues*
Strategy (of agent x)	$s_x : \Theta_x \times A \to A_x$	s_x^δ (Defn 5.3)	s_x^S (Defn 6.2)
Social choice function	$f : \Theta_\alpha \times \Theta_\beta \to \mathcal{O}$	-	f_d & f_n (Defn 6.5)

Table 3: Dialogues and dialogue sequences as mechanisms

[14] present a work on applying mechanism design in persuasion dialogues. Ours work models generic games with ABA and uses a different set of utility settings to ensure truthfulness of agents.

[20] present a work on designing mechanisms for agents in persuasion and conflict resolution dialogues with the DC system [24]. They are concerned with analysing existing dialogues with game theoretic notions whereas we have studied the interplay between dialogues and games in both directions.

[7] also use the Mackenzie dialogue system to study game strategies. The difference is that they aim at implementing dialogue protocols with game theoretic notions whereas ours is about enforcing good agents' behaviours. Argumentation and dialogue settings used in these two works are also different.

[1] present a work on studying agent strategies in persuasion dialogues. Their approach derives dialogue strategies from predefined agent profiles, e.g. *agreeable* agents that accept everything, *argumentative* agents that challenge everything, etc, but without linking dialogue results with agents' internal beliefs.

Argumentation dialogues have been studied by various researchers (e.g, see [26, 32, 5]). The dialogue model used here uses elements from the model presented in [12] and [15]. They focus on presenting the dialogue model rather than being concerned with games.

[22] present a study on collaborative agent behaviours for resource sharing with a game theoretic approach with specific constraints. It is not linked to argumentation or dialogues.

[33] extends AA to a Game-based Argumentation Framework to model agents with private knowledge engaging in sequences of argumentation dialogues, where each dialogue is associated with payoffs. That work does not formally study the relations between extensive games and argumentation dialogues as it does not model game-theoretic solution concepts with argumentation notions.

[38] investigate how to determine optimal strategies in argumentation dialogues. The authors model dialogues as extensive games with perfect information. They present computation of agents' utilities in games with different dialogue strategies. [16] use an example in [38] on modelling a legal argumentation dialogue with an extensive game. It also observes that argumentation dialogues are games of incomplete information, i.e. agents types, actions and payoffs may be private. They have not studied ensuring agents' good behaviour with games.

[18] introduce a persuasion game in which two arguers exchange arguments for the purpose of persuading an audience witnessing the argument. They study the problem in the case where there is uncertainty on audience types. Although they introduce game-theoretic elements, such as agents' payoffs, their work does not focus on modelling dialogues as games.

[17] and [25] study the computational complexity of "strategic argumentation". In both works, the authors consider argumentation dialogues not unlike ours. [17] have shown that the problem of identifying most suitable utterances in dialogues is NP-complete. They have used two argumentation formalisms, argumentation logic and agent logic, both structured argumentation formalisms, in their work. Their dialogue model also supports information exchange in the level of sub-argument, i.e. literals and rules. [25] have shown several additional complexity results with strategic argumentation. Both works focus on complexity rather than agent behaviours in the context of games.

[5] present a study on dialogue systems that support deliberation dialogues. Their underlying argumentation framework is the *instantiated value-based argumentation framework*. Their dialogue model and results are concerned with agents with preferences. Their system relies upon agents estimating their counterparts' preferences and does not study strategies with mechanism design.

8. CONCLUSION AND FUTURE WORK

We have shown how argumentation in the form of ABA can be used to solve game theory problems, both statically and distributedly via dialogues. We have built a correspondence between dominant solutions, Nash equilibria, admissible arguments, and successful dialogues. We have also studied agents' strategic behaviours in dialogues with (reverse) game-theoretic notions. Finally, we have shown that, with specific utility settings, truthfulness of and disclosure by agents can be ensured. The approach taken in the paper consists of different "layers": a bottom layer (the game to be simulated, that we refer to as the underlying game), a middle layer (the ABA framework and the dialogue sharing parts of the framework), and a top-layer (a meta-level game, on top of the first two layers, that has as a dominant solution a truthful dialogue).

This work opens many questions for future work. We have focused on dominance and Nash equilibria as solution concepts in the underlying games. Could other concepts, such as Pareto Optimality, be usefully computed via argumentation? We have considered pure strategies, what about mixed strategies in games? Could they be computed with argumentation? Also, we have considered (underlying) games in normal form. Could games in extensive form of incomplete information be represented in ABA and ABA dialogues? We have treated the meta-level dialogical game as a normal form game too, could this be interpreted instead as a game in extensive form? Results presented in this work rely upon assigning the honesty reward $\mathcal{W}$ to agents. How should $\mathcal{W}$ be estimated in general? Lemma 5.1 shows that no lower $\mathcal{W}$ could work for the given utility setting. Is there any other way of ensuring honesty under different utility settings? Is there any relation between our honesty reward and Vickrey auctions [23], in which the winner pays the second-highest bid? Finally, it would be interesting to study the computational complexity of our approach, making use of existing results for ABA [8] and instantiating our approach within practical MAS applications of games.

Acknowledgements

This research was supported by the EPSRC TRaDAr project *Transparent Rational Decisions by Argumentation*: EP/J020915/1 and by the National Research Foundation Singapore under its Interactive Digital Media (IDM) Strategic Research Programme.

REFERENCES

[1] L. Amgoud and S. Parsons. An argumentation framework for merging conflicting knowledge bases. In *JELIA Proc.*, pages 27–37, London, UK, 2002. Springer.

[2] L. Amgoud, S. Parsons, and N. Maudet. Modelling dialogues using argumentation. In *ICMAS Proc.*, pages 31–38, Boston, MA, 2000. IEEE Press.

[3] T. J. M. Bench-Capon and P. E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007.

[4] P. Besnard and A. Hunter. *Elements of Argumentation*. The MIT Press, 2008.

[5] E. Black and K. Atkinson. Choosing persuasive arguments for action. In *The 10th International Conference on Autonomous Agents and Multiagent Systems*, 2011.

[6] E. Black and A. Hunter. An inquiry dialogue system. *JAAMAS*, 19:173–209, 2009.

[7] G. Boella, J. Gerbrandy, and J. Hulstijn. A flexible mechanism for dialogue design. In *Proceedings of the 11th conference on Theoretical aspects of rationality and knowledge*, pages 160–168, 2007.

[8] Y. Dimopoulos, B. Nebel, and F. Toni. On the computational complexity of assumption-based argumentation for default reasoning. *Artificial Intelligence*, 141:57–78, 2002.

[9] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *AIJ*, 77(2):321–357, 1995.

[10] P. Dung, R. Kowalski, and F. Toni. Assumption-based argumentation. In *Argumentation in Artificial Intelligence*, pages 199–218. Springer, 2009.

[11] P. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *AIJ*, 171(10–15):642–674, 2007.

[12] X. Fan and F. Toni. Assumption-based argumentation dialogues. In *Proc. IJCAI*, 2011.

[13] X. Fan and F. Toni. Agent Strategies for ABA-based Information-seeking and Inquiry Dialogues. In *Proc. ECAI*, 2012.

[14] X. Fan and F. Toni. Mechanism design for argumentation-based persuasion. In *Proc. COMMA*. IOS Press, 2012.

[15] X. Fan and F. Toni. A general framework for sound assumption-based argumentation dialogues. *Artificial Intelligence*, 216:20–54, 2014.

[16] G. Governatori, F. Olivieri, A. Rotolo, S. Scannapieco, and G. Sartor. Two faces of strategic argumentation in the law. In *JURIX proc.*, pages 81–90, 2014.

[17] G. Governatori, F. Olivieri, S. Scannapieco, A. Rotolo, and M. Cristani. Strategic argumentation is np-complete. In *Proc. ECAI*, pages 399–404, 2014.

[18] D. Grossi and W. van der Hoek. Audience-based uncertainty in abstract argument games. In *IJCAI Proc.*, 2013.

[19] M. Jackson. Mechanism theory. In U. Derigs, editor, *Optimization and Operations Research*. EOLSS Publishers, Oxford UK, 2003.

[20] M. Kacprzak, M. Dziubinski, and K. Budzynska. Strategies in dialogues: A game-theoretic approach. In *Proc. COMMA*, pages 333–344. IOS Press, 2014.

[21] E. Kok, J.-J. Meyer, H. Prakken, and G. Vreeswijk. Testing the benfits of structured argumentation in multi-agent deliberation dialogues. In *ArgMAS Proc.*, 2012.

[22] S. Kraus and O. Schechter. Strategic-negotiation for sharing a resource between two agents. *Computational Intelligence*, 19:2003, 2003.

[23] V. Krishna. *Auction Theory*. Academic Press, 2002.

[24] J. Mackenzie. Question begging in noncumulative systems. *Journal of Philosophical Logic*, 1(8), 1979.

[25] M. Maher. Complexity of exploiting privacy violations in strategic argumentation. In *PRICAI Proc.*, pages 523–535. Springer International Publishing, 2014.

[26] P. McBurney and S. Parsons. Dialogue games for agent argumentation. In *Argumentation in Artificial Intelligence*, pages 261–280. Springer, 2009.

[27] S. Modgil and H. Prakken. The *ASPIC*$^+$ framework for structured argumentation: a tutorial. *Argument & Computation*, 5(1):31–62, 2014.

[28] S. Modgil, F. Toni, F. Bex, I. Bratko, C. Chesñevar, W. Dvořák, M. Falappa, X. Fan, S. Gaggl, A. García, M. González, T. Gordon, J. Leite, M. Možina, C. Reed, G. Simari, S. Szeider, P. Torroni, and S. Woltran. The added value of argumentation. In *Agreement Technologies*, volume 8, pages 357–403. Springer, 2013.

[29] S. Pan, K. Larson, and I. Rahwan. Argumentation mechanism design for preferred semantics. In *COMMA*, pages 403–414. IOS Press, 2010.

[30] S. Parsons, P. McBurney, E. Sklar, and M. Wooldridge. On the relevance of utterances in formal inter-agent dialogues. In *Proc. AAMAS*, pages 47–62, 2007.

[31] S. Parsons, M. Wooldridge, and L. Amgoud. Properties and complexity of some formal inter-agent dialogues. *Journal of Logic and Computation*, 13:347–376, 2003.

[32] H. Prakken. Formal systems for persuasion dialogue. *Knowledge Engineering Review*, 21(2):163–188, 2006.

[33] A. D. Procaccia and J. S. Rosenschein. Extensive-form argumentation games. In *EUMAS Proc.*, pages 312–322, 2005.

[34] I. Rahwan and K. Larson. Argumentation and game theory. In *Argumentation in Artificial Intelligence*, pages 321–339. Springer, 2009.

[35] I. Rahwan and K. Larson. Logical mechanism design. *Knowledge Engineering Review*, 26(1):61–69, 2011.

[36] I. Rahwan, K. Larson, and F. A. Tohmé. A characterisation of strategy-proofness for grounded argumentation semantics. In *Proc. IJCAI*, pages 251–256, 2009.

[37] I. Rahwan and G. R. Simari. *Argumentation in Artificial Intelligence*. Springer, 2009.

[38] R. Riveret, H. Prakken, A. Rotolo, and G. Sartor. Heuristics in argumentation: A game theory investigation. In *COMMA Proc.*, pages 324–335, 2008.

[39] P. Thang, P. Dung, and N. Hung. Towards argument-based foundation for sceptical and credulous dialogue games. In *Proc. COMMA*, pages 398–409. IOS Press, 2012.

[40] F. Toni. Reasoning on the web with assumption-based argumentation. In *Reasoning Web. Semantic Technologies for Advanced Query Answering*, volume 7487, pages 370–386. Springer, 2012.

[41] F. Toni. A tutorial on assumption-based argumentation. *Argument & Computation, Special Issue: Tutorials on Structured Argumentation*, 5(1):89–117, 2014.

[42] J. N. Webb. *Game Theory*. Springer, 2007.