Dialogical Two-Agent Decision Making with Assumption-based Argumentation

Xiuyi Fan, Francesca Toni, Andrei Mocanu*, Matthew Williams
Imperial College London, United Kingdom
{x.fan09,ft,andrei.mocanu12}@imperial.ac.uk, Matthew.williams2@imperial.nhs.uk

ABSTRACT

Much research has been devoted in recent years to argumentation-based decision making. However, less attention has been given to argumentation-based decision making amongst *multiple agents*. We present a multi-agent decision framework based on Assumption-based Argumentation. In our model, agents have goals and decisions have attributes which satisfy goals. Our framework supports agents with different goals, candidate decisions, attributes and relations amongst them. Using an existing argumentation-based dialogue framework, we show how *two agents* can argue towards "good" decisions in a distributed manner. We show that, under specific conditions, "good" decisions correspond to (1) admissible arguments for the two agents and (2) claims of successful dialogues between the two agents. Thus, this work connects decision making with multi-agent argumentation and dialogues.

Categories and Subject Descriptors

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

Keywords

Collective Decision Making, Argumentation

1. INTRODUCTION

Much research has been devoted in recent years to argumentation-based decision making [2, 14, 8]. However, less attention has been given to argumentation-based decision making amongst *multiple agents*. This work focuses on scenarios where *two agents* with potentially different interests and knowledge bases want or need to take certain decisions collaboratively and in an informed manner, such that all relevant information from both agents is considered. The two agents share common criteria towards evaluating "good" decisions. To exchange information and to take decisions interactively, the two agents conduct an argumentation-based dialogue so candidate decisions, goals and attributes are discussed.

Sharing knowledge across agents is useful when information from one agent complements information from the others. However, it might be the case that agents have different information about the same topic. In this case, conflicts may arise. To resolve conflicts, we use a specific type of trustworthiness associated with agents, if available, such that the more trusted agent is believed when discrepancies occur.

Medical decision making is an application area for our approach. Consider multiple cooperating hospitals. Each hospital, represented as an agent, has its own medical records and experimental data, and is familiar with a specific body of medical research. When a hospital is presented with a patient, possibly in some complex situation, the hospital may decide to consult some hospital which specialises in this type of patients. Hence, the two hospitals conduct a dialogue to exchange information about this patient and relevant information to make the best diagnosis or treatment decision. In this case, the specialist hospital should be more trusted.

In our work, each agent is equipped with a *decision framework* that describes *decisions*, *attributes*, *goals* and relations amongst them, as in [11]. To represent the trustworthiness of an agent, we let each agent bear a public *trust score*. The two agents also agree on the same *decision function*, which specifies the underlying decision criterion for identifying "good" decisions.

We use Assumption-based Argumentation (ABA) [7] and ABA dialogues [9] to compute, communicate and explain decisions. ABA is a widely used argumentation formalism with well understood properties. With the commonly agreed decision function and the trust score of both agents, each agent represents its knowledge in a decision framework. Upon interaction, the agents exchange the relevant parts of their decision frameworks via ABA dialogues. We show that "good" decisions correspond to *successful* dialogues (a dialogue is successful iff its claim is admissible with respect to all disclosed information). Hence, in this work we establish a connection between "good" decisions in decision making with successful dialogues and admissible arguments in argumentation.

The remainder is organised as follows. Section 2 gives background. Section 3 introduces our decision making agent model. Section 4 shows the use of ABA to compute decisions. Section 5 presents our ABA dialogues for decision making. In Section 6, to illustrate and test our model, we discuss a software realisation, in JADE [3], of our two-agent decision making dialogue system for medical literature selection. Section 7 gives several variations of our model when public trust scores are not available. Section 8 discusses related work and Section 9 concludes.

2. BACKGROUND

This work relies upon Assumption-based Argumentation (ABA), ABA dialogues, and Decision Frameworks, summarised below.

Assumption-based Argumentation (ABA) *frameworks* [7] are tu-

*Also affiliated with University of Craiova, Romania

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

ples $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$ where[1]

- $\langle \mathcal{L}, \mathcal{R} \rangle$ is a deductive system, with $\mathcal{L}$ the *language* and $\mathcal{R}$ a set of *rules* of the form $\beta_0 \leftarrow \beta_1, \ldots, \beta_m (m \geq 0, \beta_i \in \mathcal{L})$;
- $\mathcal{A} \subseteq \mathcal{L}$ is a (non-empty) set, referred to as *assumptions*;
- $\mathcal{C}$ is a total mapping from $\mathcal{A}$ into $2^{\mathcal{L}} - \{\{\}\}$, where each $\beta \in \mathcal{C}(\alpha)$ is a *contrary* of α, for $\alpha \in \mathcal{A}$.

Given a rule ρ of the form $\beta_0 \leftarrow \beta_1, \ldots, \beta_m$, β_0 is referred to as the *head* (denoted $Head(\rho) = \beta_0$) and $\beta_1, \ldots, \beta_m$ as the *body* (denoted $Body(\rho) = \{\beta_1, \ldots, \beta_m\}$). We focus on *flat* ABA frameworks, where 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. Informally, following [7]:

- *an argument for (claim)* $\beta \in \mathcal{L}$ *supported by* $A \subseteq \mathcal{A}$ ($A \vdash \beta$ in short) is a (finite) tree with nodes labelled by sentences in $\mathcal{L}$ or by τ[2], the root labelled by β, leaves either τ or assumptions in A, and non-leaves β' with, as children, the elements of the body of some rule with head β';
- *an argument* $A_1 \vdash \beta_1$ attacks an argument $A_2 \vdash \beta_2$ iff β_1 is a contrary of one of the assumptions in A_2.

Attacks between (sets of) arguments in ABA correspond to attacks between sets of assumptions, where *a set of assumptions* A *attacks a set of assumptions* A' iff an argument supported by a subset of A attacks an argument supported by a subset of A'.

With argument and attack defined for a given $AF = \langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$, standard semantics can be applied in ABA [7], e.g.: *a set of assumptions is admissible* (in AF) iff it does not attack itself and it attacks all $A \subseteq \mathcal{A}$ that attack it; *an argument* $A \vdash \beta$ *is admissible (in* AF*) supported by* $A' \subseteq \mathcal{A}$ iff $A \subseteq A'$ and A' is admissible (in AF); *a sentence is admissible* (in AF) iff it is the claim of an argument that is admissible (in AF) supported by some $A \subseteq \mathcal{A}$.

ABA frameworks can be "merged" to form *joint frameworks* [10]: 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 $AF_J = AF \uplus AF' = \langle \mathcal{L}, \mathcal{R} \cup \mathcal{R}', \mathcal{A} \cup \mathcal{A}', \mathcal{C}_J \rangle$, with $\mathcal{C}_J(\alpha) = \mathcal{C}(\alpha) \cup \mathcal{C}'(\alpha)$, for all $\alpha \in \mathcal{A} \cup \mathcal{A}'$.

ABA-dialogues [9] are conducted between two agents α_1 and α_2 that can be thought of as being equipped with ABA frameworks $\langle \mathcal{L}, \mathcal{R}_1, \mathcal{A}_1, \mathcal{C}_1 \rangle$ and $\langle \mathcal{L}, \mathcal{R}_2, \mathcal{A}_2, \mathcal{C}_2 \rangle$ respectively, sharing a common language $\mathcal{L}$. An ABA-dialogue is made of *utterances* of the form $\langle \alpha_i, \alpha_j, T, C, ID \rangle$ (for $i, j = 1, 2, i \neq j$) where: C (the *content*) is one of: $claim(\beta)$ for some $\beta \in \mathcal{L}$ (a *claim*), $rl(\beta_0 \leftarrow \beta_1, \ldots, \beta_m)$ for some $\beta_0, \ldots, \beta_m \in \mathcal{L}$ (a *rule*), $asm(\alpha)$ for some $\alpha \in \mathcal{L}$ (an *assumption*), $ctr(\alpha, \beta)$ for some $\alpha, \beta \in \mathcal{L}$ (a *contrary*), a *pass* sentence $\pi \notin \mathcal{L}$; $ID \in \mathbb{N}$ (the *identifier*); $T \in \mathbb{N} \cup \{0\}$ (the *target*) such that $T < ID$.

Utterances with content other than π or $claim(\_)$ are named *regular utterances*.[3] An utterance $\langle \alpha_i, \alpha_j, T, C, ID \rangle$ is *from* α_i *to* α_j.

Given this notion of utterances, a *dialogue* $\mathcal{D}^{\alpha_i}_{\alpha_j}(\chi)$ (*between agents* α_i *and* α_j for $\chi \in \mathcal{L}$), is a finite sequence $\delta = \langle u_1, \ldots, u_n \rangle$, $n \geq 0$, where each u_l, $l = 1, \ldots, n$, is an utterance from α_i or α_j, u_1 is an utterance from α_i, and (1) the content of u_l is $claim(\chi)$ iff $l = 1$; (2) the target of pass and claim utterances in δ is 0; (3) for every $u_i = \langle \_, \_, T, \_, \_ \rangle$ with $i > 1$ and $T \neq 0$, there is a non-pass utterance $u_k = \langle \_, \_, \_, C, T \rangle$ for $k < i$; (4) no two consecutive utterances in δ are pass utterances, other than possibly the last two utterances, u_{n-1} and u_n. Intuitively, the identifier of an utterance represents the position of the utterance in a dialogue, and its target is the identifier of some earlier utterance in the dialogue. Given a dialogue $\delta = \langle u_1, \ldots, u_n \rangle$ and an utterance u, $\delta \circ u = \langle u_1, \ldots, u_n, u \rangle$.

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 $\delta\}$;
- $\mathcal{A}_\delta = \{\alpha | asm(\alpha)$ is the content of some u_i in $\delta\}$;
- $\mathcal{C}_\delta(\alpha) = \{\beta | ctr(\alpha, \beta)$ 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 (i.e. with no assumption in the head of rules and such that all assumptions have non-empty sets of contraries), 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 cases holds:

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

Properties P1) and P2) above can be enforced using *legal-move functions*, which are mappings $\lambda : \mathcal{D} \mapsto 2^{\mathcal{U}}$ (where $\mathcal{D}$ is the set of all possible dialogues and $\mathcal{U}$ is the set of all possible utterances)[4] such that, given $\delta = \langle u_1, \ldots, u_n \rangle \in \mathcal{D}$, for all $u \in \lambda(\delta)$: $\delta \circ u$ is a dialogue and if $u = \langle \_, \_, T, C, \_ \rangle$, then there exists no $i, 1 \leq i \leq n$, such that $u_i = \langle \_, \_, T, C, \_ \rangle$. We say that δ is *compatible with* λ. Thus, there is no repeated utterance to the same target in a dialogue compatible with a legal-move function. We assume that dialogues in later discussions satisfy both P1 and P2. Λ denotes the set of all legal-move functions. *Outcome functions* are introduced to verify various dialogue properties. In particular, an *exhaustive outcome function* is defined to verify that, informally speaking, no utterance allowed by a given legal-move function is left unsaid. Dialogues fulfilling P1, P2 and some exhaustive outcome functions are termed *coherent*. *Successful* dialogues are coherent and such that the claim of a successful dialogue is admissible in the ABA framework drawn from the dialogue [9].

To generate dialogues fulfilling certain agents' aims, *strategy-move* functions [10] are used. A *strategy-move function* for agent α_i ($i = 1, 2$) is a mapping $\phi : \mathcal{D} \times \Lambda \mapsto 2^{\mathcal{U}^i}$, such that, given $\lambda \in \Lambda$ and $\delta \in \mathcal{D}$: $\phi(\delta, \lambda) \subseteq \lambda(\delta)$. Given a coherent dialogue $\mathcal{D}^{\alpha_i}_{\alpha_j}(\chi) = \delta = \langle u_1, \ldots, u_n \rangle$ compatible with a legal-move function λ and a strategy-move function ϕ for α_k ($k = i, j$), if, for all $u_m = \langle \alpha_k, \_, \_, \_, \_ \rangle, 1 < m \leq n, u_m \in \phi(\langle u_1, \ldots, u_{m-1} \rangle, \lambda)$, then we say that δ is *constructed with* ϕ with respect to α_k and α_k *uses* ϕ in δ. Furthermore, if α_i and α_j both use ϕ, then we say that δ is *constructed with* ϕ. The strategy-move function we use in this work is the *thorough strategy-move* function, ϕ_h. Informally speaking, a dialogue constructed with ϕ_h contains all information that is relevant to the topic 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 [10].

Decision frameworks [11] are tuples $\langle \mathtt{D}, \mathtt{A}, \mathtt{G}, \mathtt{T_{DA}}, \mathtt{T_{GA}} \rangle$ with
a (finite) set of decisions $\mathtt{D} = \{d_1, \ldots, d_n\}, n > 0$;
a (finite) set of attributes $\mathtt{A} = \{a_1, \ldots, a_m\}, m > 0$;

[1] In standard ABA [7] contrary maps to a single sentence. We use here an equivalent generalisation of this.

[2] $\tau \notin \mathcal{L}$ represents "true" and stands for the empty body of rules.

[3] Throughout, _ stands for an anonymous variable as in Prolog.

[4] In [9], legal-move functions have co-domain $\mathcal{U}$. Our generalisation here indicates that several utterances may be allowed next.

T^1_{DA}:

	£50	£70	£200	inSK	inPic	BST
ic	1	0	0	u	u	0
ritz	0	0	1	0	1	0

T^1_{GA}:

	£50	£70	£200	inSK	inPic	BST
cheap	1	0	0	0	0	0
near	0	0	0	1	0	0
quiet	0	0	0	0	0	1

Table 1: The decision framework, F_1, of α_1 in Example 3.1.

a (finite) set of goals $G = \{g_1, \ldots, g_l\}, l > 0$;
two tables: T_{DA}, (size $n \times m$), and T_{GA}, (size $l \times m$), such that[5],

- for all $d_i \in D, a_j \in A$, $T_{DA}[d_i, a_j]$ is either:
  1, representing that d_i *has* a_j, or
  0, representing that d_i does not have a_j, or
  u, representing unknown;
- for all $g_k \in G, a_j \in A$, $T_{GA}[g_k, a_j]$ is either
  1, representing that g_k is *satisfied* by a_j, or
  0, representing that g_k is not satisfied by a_j, or
  u, representing unknown.

We use $\mathcal{DEC}$ to denote the set of all possible decisions.

Given a decision framework $DF = \langle D, A, G, T_{DA}, T_{GA}\rangle$, a decision $d_i \in D$ *meets* a goal $g_j \in G$, with respect to DF, iff there exists an attribute $a_k \in A$, such that $T_{DA}[d_i, a_k] = 1$ and $T_{GA}[g_j, a_k] = 1$. $\gamma(d) = S$, where $d \in D, S \subseteq G$, denotes the *set of goals met by d.*

3. DECISION MAKING AGENTS

In this work, we consider two agents, α_1 and α_2, making decisions collaboratively. Each of them is equipped with an *decision framework* (DF). Until Section 7, agents also have different, publicly known *trust scores*, to represent their level of trustworthiness.

DEFINITION 3.1. *A* (Decision Making) Agent *is a pair* $\langle F, T\rangle$, *where F is a decision framework and $T \in \mathbb{N}$ is a trust score.*
We use $AGT = \{\alpha_1 = \langle F_1, T_1\rangle, \alpha_2 = \langle F_2, T_2\rangle\}$ *to denote the set of agents. We assume* $T_1 > T_2$.

The following example, adapted from [14], shows two decision making agents.

EXAMPLE 3.1. *Two agents* $\alpha_1 = \langle F_1, 10\rangle$ *and* $\alpha_2 = \langle F_2, 5\rangle$ *need to decide jointly on accommodation in London. The two agents' decision frameworks are shown in Tables 1 and 2, respectively.*[6] *α_1 has decisions {ic, ritz}, attributes {£50, £70, £200, inSK, inPic, BST }, and goals {cheap, near, quiet}. α_2 has decisions {jh, ic}, attributes {£50, £70, £200, inSK, inPic }, and goals {cheap, near}. The two agents disagree on whether £70 is cheap. Also, as we assumed in Definition 3.1, α_1 is more trustworthy.*

Intuitively, agents may have different views on decisions having attributes or attributes satisfying goals. We call such differences *conflicts*. The trust scores can be used to resolve conflicts. Information provided by the "more trusted" agent (with the highest T) is preferred to information given by the "less trusted" agent. Building upon this intuition, we define the *joint decision framework*:

DEFINITION 3.2. *Given* $\alpha_1 = \langle F_1, T_1\rangle$ *and* $\alpha_2 = \langle F_2, T_2\rangle$, $F_1 = \langle D^1, A^1, G^1, T^1_{DA}, T^1_{GA}\rangle$, $F_2 = \langle D^2, A^2, G^2, T^2_{DA}, T^2_{GA}\rangle$, *and* $T_1 > T_2$, *the* joint decision framework F_J *(of α_1 and α_2) is a tuple* $\langle D^J, A^J, G^J, T^J_{DA}, T^J_{GA}\rangle$, *in which:*

- $D^J = D^1 \cup D^2$; • $A^J = A^1 \cup A^2$; • $G^J = G^1 \cup G^2$;
- $T^J_X \in \{T^J_{DA}, T^J_{GA}\}$ *is such that:*[7]

$$T^J_X[x, y] = \begin{cases} T^1_X[x, y] & \text{if } T^1_X[x, y] \text{ is 1 or 0,} \\ T^2_X[x, y] & \text{otherwise.} \end{cases}$$

EXAMPLE 3.2. *Given α_1 and α_2 in Example 3.1, the joint decision framework F_J of α_1 and α_2 is shown in Table 3. The joint decision framework is the "union" of the two agents' decision frameworks with conflicts resolved by comparing trust scores of the two agents, e.g., given that* $T_1 > T_2$, *from* T^1_{GA}[cheap, £70] = *0 and* T^2_{GA}[cheap, £70] = *1, we obtain* T^J_{GA}[cheap, £70] = *0. Note that* T^J_{DA}*[jh,BST] = u as jh is only known to α_2 and it is unknown whether jh has attribute BST or not.*

The relations between decisions, goals and attributes distributed in two different agents' decision frameworks can be characterised by the following proposition, directly from Definition 3.2.

PROPOSITION 3.1. *Given* $\alpha_1 = \langle F_1, T_1\rangle$, $\alpha_2 = \langle F_2, T_2\rangle$, $F_1 = \langle D^1, A^1, G^1, T^1_{DA}, T^1_{GA}\rangle$, $F_2 = \langle D^2, A^2, G^2, T^2_{DA}, T^2_{GA}\rangle$, $T_1 > T_2$, *let the joint decision framework be* $F_J = \langle D^J, A^J, G^J, T^J_{DA}, T^J_{GA}\rangle$, *then* $d \in D^J$ *has* $a \in A^J$ *iff*
(1) $d \in D^1$ *has* $a \in A^1$, *or*
(2) $d \in D^2$ *has* $a \in A^2$, *and it is not the case that* $d \in D^1$ *does not have* $a \in A^1$.
Similarly, $g \in G^J$ *is satisfied by* $a \in A^J$ *iff*
(1) $g \in G^1$ *is satisfied by* $a \in A^1$, *or*
(2) $g \in G^2$ *is satisfied by* $a \in A^2$, *and it is not the case that* $g \in G^1$ *is not satisfied by* $a \in A^1$.

The two agents agree on a shared standard of "good" decisions. We use *(Multi-Agent) Decision Functions* to describe this standard:

DEFINITION 3.3. *Given two agents α_1 and α_2, and their joint decision framework* $F_J = \langle D^J, A^J, G^J, T^J_{DA}, T^J_{GA}\rangle$, *a* (Multi-Agent) Decision Function *is a mapping* $\psi : AGT \times AGT \mapsto 2^{\mathcal{DEC}}$, *such that* $\psi(\alpha_1, \alpha_2) \subseteq D^J$. *For any* $d, d' \in D^J$, *if* $\gamma(d) = \gamma(d')$ *and* $d \in \psi(\alpha_1, \alpha_2)$, *then* $d' \in \psi(\alpha_1, \alpha_2)$. *We say that* $\psi(\alpha_1, \alpha_2)$ *are* selected *by α_1 and α_2 with respect to ψ.*
We use Ψ to denote the set of all Decision Functions.

Definition 3.3 defines the basis of decision functions. Given two agents, a decision function selects some decisions from the joint decision framework. Moreover, if two decisions meet the goals and one is selected, then the other is also selected.

In this paper, we focus on a particular decision function, selecting decisions meeting goals which are ever met by any decisions, as follows.

DEFINITION 3.4. *Given agents α_1 and α_2 and their joint decision framework* $F_J = \langle D^J, A^J, G^J, T^J_{DA}, T^J_{GA}\rangle$, *the* Joint Dominant

[5]We use $T[x, y]$ to represent the cell in row labelled x and column labelled y in table T.
[6]Here, inSK, inPic, and BST stand for 'in South Kensington', 'in Piccadilly' and 'Back Street', respectively.
[7]If there is no cell for $T[x, y]$, then $T[x, y] = u$.

T^2_{DA}:

	£50	£70	£200	inSK	inPic
jh	0	1	0	1	0
ic	1	0	0	1	0

T^2_{GA}:

	£50	£70	£200	inSK	inPic
cheap	1	1	0	0	0
near	0	0	0	1	0

Table 2: The decision framework, F_2, of α_2 in Example 3.1.

$\mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}$:

	£50	£70	£200	inSK	inPic	BST
ic	1	0	0	1	0	0
ritz	0	0	1	0	1	0
jh	0	1	0	1	0	u

$\mathtt{T}^{\mathtt{J}}_{\mathtt{GA}}$:

	£50	£70	£200	inSK	inPic	BST
cheap	1	0	0	0	0	0
near	0	0	0	1	0	0
quiet	0	0	0	0	0	1

Table 3: The joint decision framework, $\mathcal{F}_J$ of α_1 and α_2.

Decision Function $\psi \in \Psi$ *is such that for all* $d \in \psi(\alpha_1, \alpha_2)$*, given* $S = \gamma(d)$*, then there is no* $g' \in \mathtt{G}^{\mathtt{J}} \setminus S$ *and* $g' \in \gamma(d')$*, where* $d' \in \mathtt{D}^{\mathtt{J}} \setminus \{d\}$*. We say that such* d *is a* jointly dominant decision*, or* d *is* jointly dominant*.*

We refer to a generic dominant decision function as ψ_d*.*

The idea of dominance is that a decision d is dominant iff the set of goals met by d is not a subset of the set of goals met by any other decisions. Note that a dominant decision d does not need to meet all goals, as illustrated in the following example.

EXAMPLE 3.3. *Given agents* α_1 *and* α_2 *in Example 3.1 and the joint decision framework* $\mathcal{F}_J$ *in Example 3.2,* $\psi_d(\alpha_1, \alpha_2) = \{ic\}$*, as ic meets both cheap and near, though it does not meet quiet.*

4. COMPUTING SELECTED DECISIONS

As seen in [14, 11], ABA can be used to model the decision making process. Given two agents α_1 and α_2, their joint decision framework and a decision function, each agent can construct an ABA framework AF_z (for $i = 1, 2$), in a way such that admissible arguments in the joint ABA framework, $AF_J = AF_1 \uplus AF_2$ (see Section 2), are selected decisions by α_1 and α_2.

We introduce *(Multi-Agent) Dominant ABA Framework* to compute joint dominant decisions. Here, each agent constructs its own ABA framework from its decision framework, and then the selected decisions are computed from the union of the two agents' ABA frameworks. Formally:

DEFINITION 4.1. *Given two agents* $\alpha_1 = \langle F_1, \mathtt{T}_1 \rangle$*,* $\alpha_2 = \langle F_2\ \mathtt{T}_2 \rangle$*, let* $F_z = \langle \mathtt{D}^z, \mathtt{A}^z, \mathtt{G}^z, \mathtt{T}^z_{\mathtt{DA}}, \mathtt{T}^z_{\mathtt{GA}} \rangle$ *be the decision framework for* $\alpha_z (z = 1, 2)$*, in which* $|\mathtt{D}^z| = n$*,* $|\mathtt{A}^z| = m$ *and* $|\mathtt{G}^z| = l$*. Then, the* (Multi-Agent) Dominant ABA Framework Corresponding to F_z *(for agent* α_z*) is* $AF_z = \langle \mathcal{L}, \mathcal{R}_z, \mathcal{A}_z, \mathcal{C}_z \rangle$*, where*[8]

- $\mathcal{R}_z$ *is such that:*
  $\alpha_1 > \alpha_2 \leftarrow \in \mathcal{R}_z$ *(since we assume* $\mathtt{T}_1 > \mathtt{T}_2$*);*
  for all $d_k \in \mathtt{D}^z$*,* $isD(d_k) \leftarrow \in \mathcal{R}_z$*;*
  for all $g_j \in \mathtt{G}^z$*,* $isG(g_j) \leftarrow \in \mathcal{R}_z$*;*
  for all $a_i \in \mathtt{A}^z$*,* $isA(a_i) \leftarrow \in \mathcal{R}_z$*;*
  for $k = 1, .., n; i = 1, .., m$*,*
  if $\mathtt{T}^z_{\mathtt{DA}}[d_k, a_i] = 0$ *then* $nHasA(d_k, a_i, \alpha_z) \leftarrow \in \mathcal{R}_z$*;*
  for $j = 1, ..l; i = 1, .., m$*,*
  if $\mathtt{T}^z_{\mathtt{GA}}[g_j, a_i] = 0$ *then* $nSat(g_j, a_i, \alpha_z) \leftarrow \in \mathcal{R}_z$*;*
  $\neg hasA(D, A, P) \leftarrow nHasA(D, A, Q), Q > P \in \mathcal{R}_z$*;*
  $\neg satBy(G, A, P) \leftarrow nSat(G, A, Q), Q > P \in \mathcal{R}_z$*;*
  $hasA(D, A) \leftarrow hasA(D, A, P), isD(D), isA(A) \in \mathcal{R}_z$*;*
  $satBy(G, A) \leftarrow satBy(G, A, P), isG(G), isA(A) \in \mathcal{R}_z$*;*
  $met(D, G) \leftarrow hasA(D, A), satBy(G, A) \in \mathcal{R}_z$*;*
  $notSel(D) \leftarrow nMet(D, G), isD(D), isG(G) \in \mathcal{R}_z$*;*
  $othersMet(D, G) \leftarrow met(D', G), D \neq D' \in \mathcal{R}_z$*;*
  nothing else is in $\mathcal{R}_z$*.*
- $\mathcal{A}_z$ *is such that:*
  for $k = 1, .., n; i = 1, .., m$*,*
  if $\mathtt{T}^z_{\mathtt{DA}}[d_k, a_i] = 1$ *then* $hasA(d_k, a_i, \alpha_z) \in \mathcal{A}_z$*;*
  for $j = 1, ..l; i = 1, .., m$*,*
  if $\mathtt{T}^z_{\mathtt{GA}}[g_j, a_i] = 1$ *then* $satBy(g_j, a_i, \alpha_z) \in \mathcal{A}_z$*;*
  for all $d_k \in \mathtt{D}^z$*,* $sel(d_k) \in \mathcal{A}_z$*;*
  for all $d_k \in \mathtt{D}^z$ *and* $g_j \in \mathtt{G}^z$*,* $nMet(d_k, g_j) \in \mathcal{A}_z$*;*
  for all $d_k \in \mathtt{D}^z$ *and* $g_j \in \mathtt{G}^z$*,* $none(d_k, g_j) \in \mathcal{A}_z$*;*
  nothing else is in $\mathcal{A}_z$*.*
- $\mathcal{C}_z$ *is such that:*
  $\mathcal{C}_z(hasA(D, A, P)) = \{\neg hasA(D, A, P)\}$
  $\mathcal{C}_z(satBy(G, A, P)) = \{\neg satBy(G, A, P)\}$
  $\mathcal{C}_z(sel(D)) = \{notSel(D)\}$*;*
  $\mathcal{C}_z(nMet(D, G)) = \{met(D, G), none(D, G)\}$*;*
  $\mathcal{C}_z(none(D, G)) = \{othersMet(D, G)\}$*.*

[8]We use here schemata, for rules and contraries, using variables (in capital letters) to stand for all their possible instances as follows: D, D' range over decisions, A ranges over attributes, G ranges over goals, and P, Q range over $\{\alpha_1, \alpha_2\}$.

Definition 4.1 gives the construction of the dominant ABA framework for each agent. In this construction, rules

$$isD(d_k) \leftarrow, isG(g_j) \leftarrow \text{ and } isA(a_i) \leftarrow$$

specify decisions, goals, attributes known by α_z. Rules with heads

$$\neg hasA(D, A, P) \text{ and } \neg satBy(G, A, P)$$

and assumptions

$$hasA(d_k, a_i, \alpha_z) \text{ and } satBy(g_j, a_i, \alpha_z)$$

and their contraries encode the defeasibility of decisions having attributes and attributes satisfying goals known by α_z. Such knowledge can be overwritten by a more trusted agent. Rules with heads

$$hasA(D, A), satBy(G, A) \text{ and } met(D, G)$$

specify the condition of a decision D meeting a goal G. The rest of the ABA framework defines the dominance condition that a decision d is dominant iff the goals not met by d are not met by any other decisions.

EXAMPLE 4.1. *Given* α_1 *in Example 3.1, the multi-agent dominant ABA framework* $\langle \mathcal{L}, \mathcal{R}, \mathcal{A}, \mathcal{C} \rangle$ *corresponding to* F_1 *has:*

- $\mathcal{R}$*:*
  $\alpha_1 > \alpha_2 \leftarrow$ $\quad isD(ic) \leftarrow$ $\quad isD(ritz) \leftarrow$
  $isA(50) \leftarrow$ $\quad isA(70) \leftarrow$ $\quad isA(200) \leftarrow$
  $isA(inSK) \leftarrow$ $\quad isA(inPic) \leftarrow$ $\quad isA(BST) \leftarrow$
  $isG(cheap) \leftarrow$ $\quad isG(quiet) \leftarrow$ $\quad isG(near) \leftarrow$
  $nHasA(ic, £200, \alpha_1) \leftarrow$ $\quad nHasA(ic, BST, \alpha_1) \leftarrow$
  $nHasA(ritz, £50, \alpha_1) \leftarrow$ $\quad nHasA(ritz, £70, \alpha_1) \leftarrow$
  $nHasA(ritz, inSK, \alpha_1) \leftarrow$ $\quad nHasA(ritz, BST, \alpha_1) \leftarrow$
  $nHasA(ic, £70, \alpha_1) \leftarrow$ $\quad nSat(cheap, £70, \alpha_1) \leftarrow$
  $nSat(cheap, £200, \alpha_1) \leftarrow$ $\quad nSat(cheap, inSK, \alpha_1) \leftarrow$
  $nSat(cheap, inPic, \alpha_1) \leftarrow$ $\quad nSat(cheap, BST, \alpha_1) \leftarrow$
  $nSat(near, £50, \alpha_1) \leftarrow$ $\quad nSat(near, £70, \alpha_1) \leftarrow$
  $nSat(near, £200, \alpha_1) \leftarrow$ $\quad nSat(near, inPic, \alpha_1) \leftarrow$
  $nSat(near, BST, \alpha_1) \leftarrow$ $\quad nSat(quiet, £50, \alpha_1) \leftarrow$
  $nSat(quiet, £70, \alpha_1) \leftarrow$ $\quad nSat(quiet, £200, \alpha_1) \leftarrow$
  $nSat(quiet, inSK, \alpha_1) \leftarrow$ $\quad nSat(quiet, inPic, \alpha_1) \leftarrow$
  all instances of the rule schemata in Definition 4.1;
- $\mathcal{A}$*:*
  $nMet(ic, cheap)$ $\quad nMet(ic, near)$ $\quad nMet(ic, quiet)$
  $nMet(ritz, cheap)$ $\quad nMet(ritz, near)$ $\quad nMet(ritz, quiet)$
  $none(ic, cheap)$ $\quad none(ic, near)$ $\quad none(ic, quiet)$
  $none(ritz, cheap)$ $\quad none(ritz, near)$ $\quad none(ritz, quiet)$

$hasA(ic, \pounds 50, \alpha_1)$ $\quad$ $hasA(ritz, \pounds 200, \alpha_1)$
$hasA(ritz, inPic, \alpha_1)$ $\quad$ $satBy(quiet, BST, \alpha_1)$
$satBy(cheap, \pounds 50, \alpha_1)$ $\quad$ $satBy(near, inSK, \alpha_1)$
$sel(ic)$ $\quad$ $sel(ritz)$

• $\mathcal{C}$*: all instances of contrary schemata in Definition 4.1.*

Definition 4.1 gives the mapping from decision frameworks to ABA frameworks for each agent. It bridges computing dominant decisions with computing admissible arguments. Hence, we can use the argumentation computation to identify selected decisions:

THEOREM 4.1. *Given agents* $\alpha_1 = \langle F_1, \mathtt{T}_1\rangle$ *and* $\alpha_2 = \langle F_2, \mathtt{T}_2\rangle$, *let* $F_J = \langle \mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}, \mathtt{G}^{\mathtt{J}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}}, \rangle$ *be their joint decision framework and* AF_1, AF_2 *be the multi-agent dominant ABA frameworks corresponding to* F_1 *and* F_2*, respectively. Then, for all* $d \in \mathtt{D}^{\mathtt{J}}$, $d \in \psi_d(\alpha_1, \alpha_2)$ *iff the argument* $\{sel(d)\} \vdash sel(d)$ *belongs to an admissible set in the joint ABA framework of* AF_1 *and* AF_2.

PROOF. Let $AF_J = AF_1 \uplus AF_2 = \langle \mathcal{L}, \mathcal{R}_J, \mathcal{A}_J, \mathcal{C}_J\rangle$.

Part A. We first show if $d \in \psi_d(\alpha_1, \alpha_2)$ then $\{sel(d)\} \vdash sel(d)$ is in an admissible set. This is to show:

1. $\{sel(d)\} \vdash sel(d)$ is an argument in AF_J;
2. there is a set of arguments Δ, such that $\Delta \cup \{sel(d)\} \vdash sel(d)$ withstands all attacks;
3. $\Delta \cup \{sel(d)\} \vdash sel(d)$ does not attack itself.

(1) Since $\mathcal{A}_J = \mathcal{A}_1 \cup \mathcal{A}_2$, and for each $d_i \in \mathtt{D}_{\mathtt{k}}$, $k = 1, 2$, there is $sel(d_i)$ in $\mathcal{A}_k$, $k = 1, 2$, respectively, therefore $sel(d_i)$ is in $\mathcal{A}_J$. By definition, $\{sel(d)\} \vdash sel(d)$ is an argument.

(2) Since the contrary of $sel(d)$ is $notSel(d)$, the only rule with head $notSel(d)$ is:

$$notSel(d) \leftarrow nMet(d, G), isD(d), isG(G),$$

and $nMet(d, G)$ are assumptions, arguments for $notSel(d)$ are:

$$\{nMet(d, G)\} \vdash notSel(d) \qquad (\diamond)$$

for all $G \in \mathtt{G}^J$. To show $\{sel(d)\} \vdash sel(d)$ withstands all attacks, we check arguments attacking $\diamond$.

$nMet(d, G)$ has two contraries $met(d, G)$ and $none(d, G)$. Any argument with claim $met(d, G)$ or $none(d, G)$ attacks $\diamond$. Since rules with head $met(d, G)$ are of the form:

$$met(d, G) \leftarrow hasA(d, A), satBy(G, A),$$

and

$$hasA(D, A) \leftarrow hasA(D, A, P), isD(D), isA(A) \text{ and}$$
$$satBy(G, A) \leftarrow satBy(G, A, P), isG(G), isA(A)$$

are the rules for $hasA(D, A)$ and $satBy(G, A)$, respectively, arguments for $met(d, G)$ are of the form:

$$\{hasA(d, A, P), satBy(G, A, P')\} \vdash met(d, G). \qquad (*)$$

Since $none(d, G)$ is an assumption, arguments for $none(d, G)$ are:

$$\{none(d, G)\} \vdash none(d, G). \qquad (**)$$

Since $d \in \psi_d(\alpha_1, \alpha_2)$, by Definition 3.4, there is no g' and $d' \neq d$ such that d' meets g' but d does not. This implies that for each goal $g \in \mathtt{G}^{\mathtt{J}}$, either **(I)** there is some attribute $a \in \mathtt{A}^{\mathtt{J}}$, such that d has a and g is satisfied by a, or **(II)** there is no a such that g is satisfied by a and there exists some $d' \neq d$ has a.

By Proposition 3.1 and Definition 4.1, for any decision $d \in \mathtt{D}^{\mathtt{J}}$, attribute $a \in \mathtt{A}^{\mathtt{J}}$, and $g \in \mathtt{G}^{\mathtt{J}}$, d meets g implies that the arguments

$$\{hasA(d, a, P)\} \vdash hasA(d, a, P), \text{ and}$$
$$\{satBy(g, a, P')\} \vdash satBy(g, a, P')$$

not being attacked. Hence, if it is case **(I)** then the argument $(*)$ is not attacked and $\{sel(d)\} \vdash sel(d)$ withstands attacks.

Since the contrary of $none(d, G)$ is $othersMet(d, G)$, and the only rule with head $othersMet(d, G)$ is:

$$othersMet(d, G) \leftarrow met(d', G), d \neq d',$$

attackers of (**) are of the form:

$$\{hasA(d', A, P), satBy(G, A, P')\} \vdash othersMet(d, G).$$

However, since d is dominant, there is no g such that g is met by d' but not d. Hence, one or both $hasA(d', A, P)$ and $satBy(G, A, P')$ cannot be "proved" (either not exists or is counter-attacked). So there is no argument for $othersMet(d, G)$ and (**) is not attacked. Hence $\{sel(d)\} \vdash sel(d)$ withstands attacks towards it.

To summarise, whichever case **(I)** or **(II)** applies, $\{sel(d)\} \vdash sel(d)$ withstands attacks towards it.

(3) It is easy to see $\Delta \cup \{sel(d)\} \vdash sel(d)$ does not attack itself. Since, arguments defend $sel(d)$ are of the form (*) and (**), the claims of (*) and (**) are not in the contraries of $sel(d)$ and vice verse. So there is no attack within $\Delta \cup \{sel(d)\} \vdash sel(d)$.

Part B. We show if $\{sel(d)\} \vdash sel(d)$ is in an admissible set then $d \in \psi_d(\alpha_1, \alpha_2)$. This is to show for every $g \in \mathtt{G}^{\mathtt{J}}$, either:

(1) g is met by d, or (2) g is not met by any $d' \in \mathtt{D}^{\mathtt{J}}$.

As seen previously, arguments defend $sel(d)$ are of the form (*) and (**) (with G unified to g). Since $sel(d)$ is in an admissible set, for every $g \in \mathtt{G}^{\mathtt{J}}$, either $met(d, g)$ or $none(d, g)$ can be proved, i.e., there exists an argument A for $met(d, g)$ or $none(d, g)$, and A withstands all attacks towards it. If A is of the form (*), d meets g; if A is of the form (**), then it implies there is no $d' \in \mathtt{D}^{\mathtt{J}}$ meets g. Hence, $d \in \psi_d(\alpha_1, \alpha_2)$. □

As illustrated in [11] and [8], in addition to computing decisions, ABA is good at explaining the results of its computation. Hence, computation and explanation are unified processes in ABA.

5. DECISION MAKING DIALOGUES

Thus far, we have shown how "good" decisions can be identified by computing admissible arguments in the joint framework of the two agents' ABA frameworks. The remaining problem is to implement the computation interactively. We use dialogical argumentation, and ABA dialogues in particular, to realise this goal.

The dialogue model in [9] is sound, such that given a *coherent* dialogue (see Section 2), the claim of a successful dialogue corresponds to an admissible argument in the ABA framework drawn from the dialogue. [10] extends this result and defines *strategy-move* functions for linking claims of successful dialogues with admissible arguments in the joint framework of the two agents. Results in [10] give a form of completeness result for the dialogue model, such that if a sentence is admissible in the joint framework of the two agents, then there is a successful dialogue for this sentence. Here, these results can be applied for decision making dialogues.

Since the aim of the two agents is to jointly make informed decisions, they will be truthful and disclose all relevant information. Hence, they will both implement a *thorough strategy-move* function ϕ_h while constructing dialogues (see Section 2). To show that dialogues can be used to reach decisions, we first show termination.

PROPOSITION 5.1. *Given two agents α_1, α_2, let the joint decision framework be $\langle \mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}, \mathtt{G}^{\mathtt{J}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}} \rangle$, and let $\delta = \mathcal{D}^{\alpha_i}_{\alpha_j}(sel(d))$ ($\alpha_i, \alpha_j \in \{\alpha_1, \alpha_2\}, \alpha_i \neq \alpha_j$), for any $d \in \mathtt{D}^{\mathtt{J}}$. If δ is constructed with ϕ_h, then δ terminates.*

PROOF. Since δ is constructed with ϕ_h, all contents of utterances in δ come from two agent's ABA frameworks. Since the two agents' ABA frameworks are obtained from their (finite) decision frameworks, these two ABA frameworks are finite. Also, by Definition 5 in [9], agents will not make repeated utterances to the same target. Hence, to show δ terminates is to show there is no cycle in:

1. constructing arguments for or against $sel(d)$ (and arguments for or against the attackers or defenders), and
2. attacking or defending arguments for or against $sel(d)$.

(1) can be shown by examining all rules in Definition 4.1, there is no "infinite argument" constructed from the two ABA frameworks, i.e., there is no circular rules such as $p \leftarrow q$ and $q \leftarrow p$ coexist. (2) can be shown by examining all assumptions and contraries in Definition 4.1. Similarly, there is no circular attacks. Therefore, the constructed dialogue is guaranteed to be finite. □

We illustrate a decision making dialogue as follows.

EXAMPLE 5.1. *(Example 4.1, continued.) Given the two agents shown in Example 3.1, part of a successful dialogue $\mathcal{D}^{\alpha_1}_{\alpha_2}(sel(d))$ constructed with ϕ_h is shown in Table 4. In this table, we see the initial 20 utterances in which the two agents discuss the decision ic meeting the goal cheap. They verify that ic costs £50 and cheap is satisfied by £50 by letting α_2 put forward the two assumptions $hA(ic, £50, \alpha_2)$ and $sB(cp, £50, \alpha_2)$ and receiving no objections from α_1. Not shown in the table, the two agents subsequently confirm that ic meets near and no decision meets quiet. Hence ic is a dominant decision meeting goals cheap and near.*

With results accumulated so far, we are ready to show the correspondence between joint decisions and dialogues, as follows.

THEOREM 5.1. *Given agents α_1, α_2, let the joint decision framework be $\langle \mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}, \mathtt{G}^{\mathtt{J}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}} \rangle$. Then, for any $d \in \mathtt{D}^{\mathtt{J}}$, d is jointly dominant iff there exists a successful $\delta = \mathcal{D}^{\alpha_i}_{\alpha_j}(sel(d))$ ($\alpha_i, \alpha_j \in \{\alpha_1, \alpha_2\}$, $\alpha_i \neq \alpha_j$) constructed with ϕ_h.*

PROOF. Let AF_J be the joint ABA framework of α_1 and α_2.

We first show if δ is successful, then d is jointly dominant. By Theorem 4.1, we know that d is jointly dominant iff $A = \{sel(d)\} \vdash sel(d)$ is in an admissible set in AF_J. By Theorem 1 in [10], we know that A is admissible in AF_J iff A is admissible in $\mathcal{F}_\delta$, for $\delta = \mathcal{D}^{\alpha_i}_{\alpha_j}(sel(d))$ constructed with ϕ_h. Therefore, if δ is successful, then d is jointly dominant.

We then show if d is jointly dominant, then there exists a successful dialogue. Since d is jointly dominant, then A is admissible in AF_J. By Proposition 5.1, there exists a finitely constructed δ for $sel(d)$, such that A is admissible in $\mathcal{F}_\delta$. By the definition of a successful dialogue, such δ constructed above is successful. □

6. EXPERIMENTS

To validate our approach, we have implemented a two-agent decision making dialogue system using JADE [3] and Grapharg [6] for implementing legal-move and outcome functions to realise thorough dialogues. We have used the medical literature data reported in [8], which contains 11 papers, each with 10 attributes. Table 5 shows a fragment of this data. The two agents represent two hospitals. Together, the agents want to identify the most relevant medical paper for a patient of interest. We have experimented with two hypothetical patients, with the following characteristics:

Table 4: Dialogue for the two agents (see Example 5.1). Here, s, nS, m, nM, hA, nHA, sB, nSB stand for sel, notSel, met, notMet, hasA, nHasA, satBy, nSat, respectively.

$\langle \alpha_1, \alpha_2, 0, claim(s(ic)), 1\rangle$
$\langle \alpha_2, \alpha_1, 1, asm(s(ic)), 2\rangle$
$\langle \alpha_1, \alpha_2, 2, ctr(s(ic), nS(ic)), 3\rangle$
$\langle \alpha_2, \alpha_1, 3, rl(nS(ic) \leftarrow nM(ic, cp), isD(ic), isG(cp)), 4\rangle$
$\langle \alpha_1, \alpha_2, 4, asm(nM(ic, cp)), 5\rangle$
$\langle \alpha_2, \alpha_1, 4, rl(isD(ic) \leftarrow), 6\rangle$
$\langle \alpha_1, \alpha_2, 4, rl(isG(ic) \leftarrow), 7\rangle$
$\langle \alpha_1, \alpha_2, 5, ctr(nM(ic, cp), m(ic, cp), none(ic, cp)), 8\rangle$
$\langle \alpha_1, \alpha_2, 8, rl(m(ic, cp) \leftarrow hA(ic, £50), sB(cp, £50)), 9\rangle$
$\langle \alpha_2, \alpha_1, 9, rl(hA(ic, £50) \leftarrow hA(ic, £50, \alpha_2)), 10\rangle$
$\langle \alpha_2, \alpha_1, 10, asm(hA(ic, £50, \alpha_2)), 11\rangle$
$\langle \alpha_2, \alpha_1, 9, rl(sB(cp, £50) \leftarrow sB(cp, £50, \alpha_2)), 12\rangle$
$\langle \alpha_2, \alpha_1, 12, asm(sB(cp, £50, \alpha_2), 13\rangle$
$\langle \alpha_2, \alpha_1, 11, ctr(hA(ic, £50, \alpha_2, \neg hA(ic, £50, \alpha_2), 14\rangle$
$\langle \alpha_2, \alpha_1, 12, ctr(sB(cp, £50, \alpha_2, \neg sB(cp, £50, \alpha_2), 15\rangle$
$\langle \alpha_2, \alpha_1, 14,$
$rl(\neg hA(ic, £50, \alpha_2) \leftarrow nHA(ic, £50, \alpha_1), \alpha_1 > \alpha_2), 16\rangle$
$\langle \alpha_1, \alpha_2, 16, rl(\alpha_1 > \alpha_2 \leftarrow), 17\rangle$
$\langle \alpha_1, \alpha_2, 15,$
$rl(\neg sB(cp, £50, \alpha_2) \leftarrow nSB(cp, £50, \alpha_1), \alpha_1 > \alpha_2), 18\rangle$
$\langle \alpha_1, \alpha_2, 18, rl(\alpha_1 > \alpha_2 \leftarrow), 19\rangle$
$\langle \alpha_2, \alpha_1, 3, rl(nS(ic) \leftarrow nM(ic, nr), isD(ic), isG(nr)), 20\rangle$
...

	$> 18Years$	...	$> 2metastases$	$PS\,0,1$	...
p_1	1		1	1	
p_2				1	
...					

Table 5: A fragment of the Paper / Trial Characteristics data (PS means Performance Score).

- patient1: 88 years old with lung cancer, extra-cranial diseases, a performance score of 3 and 2 brain metastases.
- patient2: 16 years old with lung cancer, extra-cranial diseases, a performance score of 2 and 1 brain matastasis.

In addition to the data reported in [8], we have also randomly generated some medical literature data, giving a total of 31 (2 to 32) papers, each with the same 10 attributes reported in [8], for experiments. Table 6 shows the number of utterances made by both agents in successful dialogues and the total execution time of the platform for a number of runs, with each run using a different number of papers as data set. We see that the run time is proportional to the number of papers.

7. GENERALISATION

Thus far, we have assumed that the two agents have different publicly known trust scores. However, such assumption might not be true as trust scores may be unavailable. Nevertheless, conflicts still need to be resolved. In this section, we give three new ways of constructing joint decision frameworks without trust scores, representing different ways of resolving conflicts.

DEFINITION 7.1. *Given two agents $\langle F_1, \_\rangle$ and $\langle F_2, \_\rangle$, $F_1 = \langle \mathtt{D}^1, \mathtt{A}^1, \mathtt{G}^1, \mathtt{T}^1_{\mathtt{DA}}, \mathtt{T}^1_{\mathtt{GA}} \rangle$, $F_2 = \langle \mathtt{D}^2, \mathtt{A}^2, \mathtt{G}^2, \mathtt{T}^2_{\mathtt{DA}}, \mathtt{T}^2_{\mathtt{GA}} \rangle$, the* sceptical joint decision framework *F_J is a tuple $\langle \mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}, \mathtt{G}^{\mathtt{J}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}} \rangle$:*

# Papers	# Utterances		Run Time (ms)	
	Patient 1	Patient 2	Patient 1	Patient 2
2	204	166	202	173
4	396	326	153	179
8	1658	691	234	198
11 [8]	1372	1258	208	198
16	2319	1789	280	207
32	6929	5514	646	494

Table 6: Experiment results with 2, 4, 8, 11, 16, and 32 papers.

- $\mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}$ *and* $\mathtt{G}^{\mathtt{J}}$ *are as given in Definition 3.2.*
- $\mathtt{T}^{\mathtt{J}}_{\mathtt{X}} \in \{\mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}}\}$ *is such that:*

$$\mathtt{T}^{\mathtt{J}}_{\mathtt{X}}[x,y] = \begin{cases} 1 & \text{if } \mathtt{T}^{1}_{\mathtt{X}}[x,y] = \mathtt{T}^{2}_{\mathtt{X}}[x,y] = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 7.1 is sceptical towards conflicts. A decision d has an attribute a or a goal g is satisfied by a iff both agents believe so.

Definition 7.2. *Given* $\langle F_1, \_\rangle$ *and* $\langle F_2, \_\rangle$, $F_1 = \langle \mathtt{D}^1, \mathtt{A}^1, \mathtt{G}^1, \mathtt{T}^1_{\mathtt{DA}}, \mathtt{T}^1_{\mathtt{GA}}\rangle$, $F_2 = \langle \mathtt{D}^2, \mathtt{A}^2, \mathtt{G}^2, \mathtt{T}^2_{\mathtt{DA}}, \mathtt{T}^2_{\mathtt{GA}}\rangle$, *the* credulous joint decision framework F_J *is a tuple* $\langle \mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}, \mathtt{G}^{\mathtt{J}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}}\rangle$*:*
- $\mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}$ *and* $\mathtt{G}^{\mathtt{J}}$ *are as given in Definition 3.2.*
- $\mathtt{T}^{\mathtt{J}}_{\mathtt{X}} \in \{\mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}}\}$ *is such that:*

$$\mathtt{T}^{\mathtt{J}}_{\mathtt{X}}[x,y] = \begin{cases} 1 & \text{if } \mathtt{T}^{1}_{\mathtt{X}}[x,y] = 1 \text{ or } \mathtt{T}^{2}_{\mathtt{X}}[x,y] = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 7.2 is credulous towards conflicts. A decision d has an attribute a or a goal g is satisfied by a iff one of the agents believes so, regardless of what the other agent thinks.

Definition 7.3. *Given* $\langle F_1, \_\rangle$ *and* $\langle F_2, \_\rangle$, $F_1 = \langle \mathtt{D}^1, \mathtt{A}^1, \mathtt{G}^1, \mathtt{T}^1_{\mathtt{DA}}, \mathtt{T}^1_{\mathtt{GA}}\rangle$, $F_2 = \langle \mathtt{D}^2, \mathtt{A}^2, \mathtt{G}^2, \mathtt{T}^2_{\mathtt{DA}}, \mathtt{T}^2_{\mathtt{GA}}\rangle$, *the* fair joint decision framework F_J *is a tuple* $\langle \mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}, \mathtt{G}^{\mathtt{J}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}}\rangle$*:*
- $\mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}$ *and* $\mathtt{G}^{\mathtt{J}}$ *are as given in Definition 3.2.*
- $\mathtt{T}^{\mathtt{J}}_{\mathtt{X}} \in \{\mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}}\}$ *is such that, for* $i, j \in \{0,1\}, i \neq j$*:*

$$\mathtt{T}^{\mathtt{J}}_{\mathtt{X}}[x,y] = \begin{cases} 1 & \text{if } \mathtt{T}^{\mathtt{i}}_{\mathtt{X}}[x,y] = 1 \text{ and } \mathtt{T}^{\mathtt{j}}_{\mathtt{X}}[x,y] \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 7.3 gives a balance between the two views above. It specifies that a decision d has an attribute a or a goal g is satisfied by a iff one agent believes so and the other agent either agrees or has no knowledge to object.

As shown previously, ABA frameworks can be constructed to compute, communicate and explain decisions. We give the ABA framework construction for making dominant decisions with sceptical / credulous / fair joint decision frameworks below.

Definition 7.4. *Given* $\alpha_1 = \langle F_1, \_\rangle$, $\alpha_2 = \langle F_2\ \_\rangle$, *let* $F_z = \langle \mathtt{D}^z, \mathtt{A}^z, \mathtt{G}^z, \mathtt{T}^z_{\mathtt{DA}}, \mathtt{T}^z_{\mathtt{GA}}\rangle$ *be the decision framework for* $\alpha_z \in \{\alpha_1, \alpha_2\}$, *in which* $|\mathtt{D}^z| = n$, $|\mathtt{A}^z| = m$ *and* $|\mathtt{G}^z| = l$, *the* multi-agent sceptical dominant ABA framework corresponding to F_z *(for agent* α_z*) is* $AF_z = \langle \mathcal{L}, \mathcal{R}_z, \mathcal{A}_z, \mathcal{C}_z\rangle$, *where*

- $\mathcal{R}_z$ *is such that:*
  for all $d_k \in \mathtt{D}^z$, $isD(d_k) \leftarrow \in \mathcal{R}_z$*;*
  for all $g_j \in \mathtt{G}^z$, $isG(g_j) \leftarrow \in \mathcal{R}_z$*;*
  for all $a_i \in \mathtt{A}^z$, $isA(a_i) \leftarrow \in \mathcal{R}_z$*;*
  for $k = 1,..,n; j = 1,..,m$,
  if $\mathtt{T}^z_{\mathtt{DA}}[d_k, a_i] = 1$ *then* $hasA(d_k, a_i, \alpha_z) \leftarrow \in \mathcal{R}_z$*;*
  for $j = 1,..m; i = 1,..,l$,
  if $\mathtt{T}^z_{\mathtt{GA}}[g_j, a_i] = 1$ *then* $satBy(g_j, a_i, \alpha_z) \leftarrow \in \mathcal{R}_z$*;*
  $hasA(D,A) \leftarrow hasA(D,A,P), isD(D), isA(A) \in \mathcal{R}_z$*;*
  $satBy(G,A) \leftarrow satBy(G,A,P), isG(G), isA(A) \in \mathcal{R}_z$*;*
  $met(D,G) \leftarrow hasA(D,A), satBy(G,A) \in \mathcal{R}_z$*;*
  $notSel(D) \leftarrow nMet(D,G), isD(D), isG(G) \in \mathcal{R}_z$*;*
  $othersMet(D,G) \leftarrow met(D',G), D \neq D' \in \mathcal{R}_z$*;*
  nothing else is in $\mathcal{R}_z$.
- $\mathcal{A}_z$ *is such that:*
  for all $d_k \in \mathtt{D}$, $sel(d_k) \in \mathcal{A}_z$*;*
  for all $d_k \in \mathtt{D}$ *and* $g_j \in \mathtt{G}$, $nMet(d_k, g_j) \in \mathcal{A}_z$*;*
  for all $d_k \in \mathtt{D}$ *and* $g_j \in \mathtt{G}$, $none(d_k, g_j) \in \mathcal{A}_z$*;*
  nothing else is in $\mathcal{A}_z$.
- $\mathcal{C}_z$ *is such that:*
  $\mathcal{C}_z(sel(D)) = \{notSel(D)\}$*;*
  $\mathcal{C}_z(nMet(D,G)) = \{met(D,G), none(D,G)\}$*;*
  $\mathcal{C}_z(none(D,G)) = \{othersMet(D,G)\}$.

Compared with Definition 4.1, Definition 7.4 introduces rules

$$hasA(d_k, a_i, \alpha_z) \leftarrow \text{ and } satBy(g_j, a_i, \alpha_z) \leftarrow$$

to replace assumptions $hasA(d_k, a_i, \alpha_z)$ and $satBy(g_j, a_i, \alpha_z)$. Hence decisions having attributes and attributes satisfying goals are no longer defeasible.

Definition 7.5. *Given* $\alpha_1 = \langle F_1, \_\rangle$, $\alpha_2 = \langle F_2\ \_\rangle$, *the* multi-agent credulous dominant ABA framework corresponding to F_z *is* AF_z*, as given in Definition 7.4, except the rules*

$$hasA(D,A) \leftarrow hasA(D,A,P), isD(D), isA(A), \text{ and}$$
$$satBy(G,A) \leftarrow satBy(G,A,P), isG(G), isA(A)$$

are, respectively, replaced by

$$hasA(D,A) \leftarrow hasA(D,A,P), hasA(D,A,Q), isD(D), isA(A), P \neq Q, \text{ and}$$
$$satBy(G,A) \leftarrow satBy(G,A,P), satBy(G,A,Q), isG(G), isA(A), P \neq Q.$$

As indicated by the changes, when constructing ABA frameworks for credulous joint decision frameworks, to have decisions having attributes or attributes satisfying assumption, both agents (represented by P and Q) need to agree on the given D, G and A.

Definition 7.6. *Given* $\alpha_1 = \langle F_1, \_\rangle$, $\alpha_2 = \langle F_2\ \_\rangle$, *the* multi-agent fair dominant ABA framework *is* AF_z*, as given in Definition 4.1, with the following changes:*
replace rules:

$$\neg hasA(D,A,P) \leftarrow nHasA(D,A,Q), Q > P \text{ and}$$
$$\neg satBy(G,A,P) \leftarrow nSat(G,A,Q), Q > P$$

with

$$\neg hasA(D,A,P) \leftarrow nHasA(D,A,Q) \text{ and}$$
$$\neg satBy(G,A,P) \leftarrow nSat(G,A,Q).$$

Remove $P > Q \leftarrow$ *from* $\mathcal{R}_z$ *for* $P, Q \in \{\alpha_1, \alpha_2\}$.

The intuition here is that we drop the trust score comparison when checking if the agent has conflicting information. Hence, a decision d does not have an attribute a if one of the agent believes d does not have a. The same holds for attributes satisfying goals. These encode the behavior of fair joint decision frameworks.

With newly defined sceptical, credulous and fair joint decision frameworks, our earlier results that "good" decisions are admissible arguments still hold, as shown in the following theorem. (We omit the proof due to the lack of space, though it is very similar to the one given in Theorem 4.1.)

THEOREM 7.1. *Given* $\alpha_1 = \langle F_1, \_\rangle$ *and* $\alpha_2 = \langle F_2, \_\rangle$, *let* $F_J = \langle \mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}, \mathtt{G}^{\mathtt{J}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}}, \rangle$ *be the sceptical (credulous / fair) joint decision framework and* AF_1, AF_2 *be the multi-agent sceptical (credulous / fair) dominant ABA framework corresponding to* F_1 *and* F_2, *respectively. Then, for all* $d \in \mathtt{D}^{\mathtt{J}}$, $d \in \psi_d(\alpha_1, \alpha_2)$ *iff* $\{sel(d)\} \vdash sel(d)$ *belongs to an admissible set in* $AF_1 \uplus AF_2$.

Similarly, the results about dialogues hold as well, shown below. (The proof is omitted but uses Theorem 7.1 and Theorem 1 in [10].)

THEOREM 7.2. *Given* α_1, α_2, *let the sceptical (credulous / fair) joint decision framework be* $\langle \mathtt{D}^{\mathtt{J}}, \mathtt{A}^{\mathtt{J}}, \mathtt{G}^{\mathtt{J}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{DA}}, \mathtt{T}^{\mathtt{J}}_{\mathtt{GA}} \rangle$. *Then for any* $d \in \mathtt{D}^{\mathtt{J}}$, d *is jointly dominant iff there exists a successful* $\mathcal{D}^{\alpha_i}_{\alpha_j}(sel(d))$ ($\alpha_i, \alpha_j \in \{\alpha_1, \alpha_2\}$, $\alpha_i \neq \alpha_j$) *constructed with* ϕ_h.

8. RELATED WORK

Matt et.al. [14], Fan and Toni [11], and Fan et.al. [8] present several argumentation-based decision-making models. Though they all use ABA and introduce decision functions, our work differs from theirs in that we study decision making with two agents using dialogues whereas they have been focusing on single agent decision making. Also, our decision certeria are different, e.g., trust scores are used for resolving conflicts between agents.

Amgoud and Prade [2] present a formal model for making decisions using abstract argumentation. Our work differs from theirs as: (1) they use abstract argumentation and a specialised instantiation whereas we use ABA; (2) they have not considered decision making in the context of multiple agents or considered dialogues.

Amgoud et.al. [1] present a work on consensus forming. They propose a dialogical decision making model in which arguments for and against decisions are presented. Our work differs from theirs as (1) we use ABA whereas they use a propositional language; (2) decision making in our work has been given a normative characterisation in the way that "good" decisions are defined without refering to arguments, whereas they require a tight coupling between arguments and decisions; (3) our dialogue model is sound and complete, whereas they have presented a protocol with no formal results.

Kakas and Moraïtis [12] present a decision making model focused on modelling agent personality. The differences are: (1) we use ABA whereas they use logic programming; (2) our work has the aforementioned normative characterisation whereas theirs has not; (3) our work connects decision making with dialogues.

Müller and Hunter [16] present a decision making model based on a simplified version of ASPIC+ [17]. Our work is similar to theirs as both have a normative characterisation though with different decision criteria. They have focused on a single agent decision making without dialogues whereas our work is two-agent decision making with dialogues.

[5, 13, 15, 4] present dialogue models on inquiry and deliberation. Our work gives a formal decision model whereas theirs are solely concerned with dialogues.

9. CONCLUSION AND FUTURE WORK

In this work, we presented a two-agent decision making model with argumentation dialogues. The aim of the two agents is to make informed decisions by using information from both agents. In our model, the two agents carry possibly different decision frameworks. The two agents share a common standard for "good" decisions. While making decisions, each agent constructs an ABA framework corresponding to its decision framework. Then an ABA dialogue is conducted between the two agents. We give several ways of resolving conflicts between agents, with or without trust measures. We show that under the condition that both agents are truthful and disclose all relevant information, successful dialogues identify "good" decisions. Although simple, our approach has wide applicabilities, e.g., in the medical domain as we illustrated. The main contribution of our work is that it performs a "complete cycle", from decision frameworks, to argumentation-based computation, to dialogues, to a real world application with implementation. No similar effort has been reported in the literature before, to the best of our knowledge.

Future work lies in at least four directions. Firstly, we would like to study decision making with agent preferences, i.e., agents prefer certain goals over others. Secondly, we would like to study decision making with more flexible knowledge representation such that information is not in well-formatted tables. This involves arguments about what is believed to be the case, about attributes of decision options, etc. Thirdly, we would like to study decision making in a game theoretical setting in which agents are not completely cooperative. Lastly, we would like to expand our dialogue model so more agents can be supported.

Acknowledgements. This research was partially supported by the EPSRC project *Transparent Rational Decisions by Argumentation*: EP/J020915/1.

10. REFERENCES

[1] L. Amgoud, S. Belabbes, and H. Prade. Towards a formal framework for the search of a consensus between autonomous agents. In *Proc. AAMAS*, 2005.

[2] L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3-4), 2009.

[3] F. Bellifemine, A. Poggi, and G. Rimassa. JADE - a FIPA-compliant agent framework. In *Proc. PAIA*, 1999.

[4] E. Black and K. Atkinson. Choosing persuasive arguments for action. In *Proc. of AAMAS*, 2011.

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

[6] R. Craven, F. Toni, and M. Williams. Graph-based dispute derivations in assumption-based argumentation. In *Proc. TAFA*, 2013.

[7] P. M. Dung, R. A. Kowalski, and F. Toni. Assumption-based argumentation. In *Arg. in AI*, pages 25–44. Springer, 2009.

[8] X. Fan, R. Craven, R. Singer, F. Toni, and M. Williams. Assumption-based argumentation for decision-making with preferences: A medical case study. In *Proc. CLIMA*, 2013.

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

[10] X. Fan and F. Toni. Agent strategies for ABA-based information-seeking and inquiry dialogues. In *Proc. ECAI*, 2012.

[11] X. Fan and F. Toni. Decision making with assumption-based argumentation. In *Proc. TAFA*, 2013.

[12] A. Kakas and P. Moraitis. Argumentation based decision making for autonomous agents. In *Proc. AAMAS*, 2003.

[13] E. M. Kok, J.-J. C. Meyer, H. Prakken, and G. Vreeswijk. A formal argumentation framework for deliberation dialogues. In *Proc. ArgMAS*, 2010.

[14] P. A. Matt, F. Toni, and J. Vaccari. Dominant decisions by argumentation agents. In *Proc. ArgMAS*. Springer, 2009.

[15] P. McBurney, D. Hitchcock, and S. Parsons. The eightfold way of deliberation dialogue. *IJIS*, 22(1):95–132, 2007.

[16] J. Müller and A. Hunter. An argumentation-based approach for decision making. In *Proc. ICTAI*, 2012.

[17] H. Prakken. An abstract framework for argumentation with structured arguments. *Arg. and Comp.*, 1(2):93–124, 2010.