Agreeing what to do

Elizabeth Black[1] and Katie Atkinson[2]

[1] Department of Engineering Science, University of Oxford, UK
lizblack@robots.ox.ac.uk
[2] Department of Computer Science, University of Liverpool, UK
katie@liverpool.ac.uk

Abstract. When deliberating about what to do, an autonomous agent must generate and consider the relative pros and cons of the different options. The situation becomes even more complicated when an agent is involved in a joint deliberation, as each agent will have its own preferred outcome which may change as new information is received from the other agents involved in the deliberation. We present an argumentation-based dialogue system that allows agents to come to an agreement on how to act in order to achieve a joint goal. The dialogue strategy that we define ensures that any agreement reached is acceptable to each agent, but does not necessarily demand that the agents resolve or share their differing preferences. We give properties of our system and discuss possible extensions.

ACM Category: I.2.11 Multiagent systems. **General terms:** Theory.
Keywords: dialogue, argumentation, agreement, strategy, deliberation, action.

1 Introduction

When agents engage in dialogues their behaviour is influenced by a number of factors including the type of dialogue taking place (e.g. negotiation or inquiry), the agents' own interests within the dialogue, and the other parties participating in the dialogue. Some of these aspects have been recognised in Walton and Krabbe's characterisation of dialogue types [1]. Some types of dialogue are more adversarial than others. For example, in a persuasion dialogue an agent may try to force its opponent to contradict itself, thus weakening the opponent's position. In a deliberation dialogue, however, the agents are more co-operative as they each share the same goal to establish agreement, although individually they may wish to influence the outcome in their own favour.

We present a dialogue system for deliberation that allows agents to reason and argue about what to do to achieve some joint goal but does not require them to pool their knowledge, nor does it require them to aggregate their preferences. Few existing dialogue systems address the problem of deliberation ([2, 3] are notable exceptions). Ours is the first system for deliberation that provides a dialogue strategy that allows agents to come to an agreement about how to act that each is happy with, despite the fact that they may have different preferences and thus may each be agreeing for different reasons; it couples a dialectical setting with formal methods for argument evaluation and allows strategic manoeuvring in order to influence the dialogue outcome. We present an analysis of when agreement can and cannot be reached with our system; this provides

an essential foundation to allow us to explore mechanisms that allow agents to come to an agreement in situations where the system presented here may fail.

We assume that agents are co-operative in that they do not mislead one another and will come to an agreement wherever possible; however, each agent aims to satisfy its own preferences. For the sake of simplicity, here we present a two party dialogue; however, the assumed co-operative setting means that many of the difficult issues which normally arise with multi party dialogues (e.g. [4]) are avoided here. We believe it to be straightforward to extend the system to allow multiple participants, for example following the approach taken in [5].

We describe the setting envisaged through a characteristic scenario. Consider a situation where a group of colleagues is attending a conference and they would all like to go out for dinner together. Inevitably, a deliberation takes place where options are proposed and critiqued and each individual will have his own preferences that he wishes to be satisfied by the group's decision. It is likely that there will be a range of different options proposed that are based on criteria such as: the type of cuisine desired; the proximity of the restaurant; the expense involved; the restaurant's capacity; etc.

To start the dialogue one party may put forward a particular proposal, reflecting his own preferences, say going to a French restaurant in the town centre. Such an argument may be attacked on numerous grounds, such as it being a taxi ride away, or it being expensive. If expense is a particular consideration for some members of the party, then alternative options would have to be proposed, each of which may have its own merits and disadvantages, and may need to consider the preferences already expressed. We can see that in such a scenario the agents , whilst each having their own preferred options, are committed to finding an outcome that everyone can agree to.

We present a formal argumentation-based dialogue system to handle joint deliberation. In section 2 we present the reasoning mechanism through which agents can construct and propose arguments about action. In section 3 we define the dialogue system and give an example dialogue. In section 4 we present an analysis of our system and in section 5 we discuss important extensions. In section 6 we discuss related work, and we conclude the paper in section 7.

2 Practical arguments

We now describe the model of argumentation that we use to allow agents to reason about how to act. Our account is based upon a popular approach to argument characterisation, whereby argumentation schemes and critical questions are used as presumptive justification for generating arguments and attacks between them [6]. Arguments are generated by an agent instantiating a *scheme for practical reasoning* which makes explicit the following elements: the initial circumstances where action is required; the action to be taken; the new circumstances that arise through acting; the goal to be achieved; and the social value promoted by realising the goal in this way. The scheme is associated with a set of characteristic critical questions (CQs) that can be used to identify challenges to proposals for action that instantiate the scheme. An unfavourable answer to a CQ will identify a potential flaw in the argument. Since the scheme makes use of what are termed as 'values', this caters for arguments based on subjective preferences as well

as more objective facts. Such values represent qualitative social interests that an agent wishes (or does not wish) to uphold by realising the goal stated [7].

To enable the practical argument scheme and critical questions approach to be precisely formalised for use in automated systems, in [8] it was defined in terms of an Action-based Alternating Transition System (AATS) [9], which is a structure for modelling game-like multi-agent systems where the agents can perform actions in order to attempt to control the system in some way. Whilst the formalisms given in [8, 9] are intended to represent the overall behaviour of a multi-agent system and the effects of joint actions performed by the agents, we are interested in representing the knowledge of individual agents within a system. Hence, we use an adaptation of their formalisms (first presented in [5]) to define a *Value-based Transition System* (VATS) as follows.

Definition 1: *A* **Value-based Transition System** *(***VATS***), for an agent x, denoted S^x, is a 9-tuple $\langle Q^x, q_0^x, Ac^x, Av^x, \rho^x, \tau^x, \Phi^x, \pi^x, \delta^x \rangle$ s.t.:*

Q^x *is a finite set of* states*;*

$q_0^x \in Q^x$ *is the designated* initial state*;*

Ac^x *is a finite set of* actions*;*

Av^x *is a finite set of* values*;*

$\rho^x : Ac^x \mapsto 2^{Q^x}$ *is an* action precondition function*, which for each action $a \in Ac^x$ defines the set of states $\rho(a)$ from which a may be executed;*

$\tau^x : Q^x \times Ac^x \mapsto Q^x$ *is a partial* system transition function*, which defines the state $\tau^x(q, a)$ that would result by the performance of a from state q—n.b. as this function is partial, not all actions are possible in all states (cf. the precondition function above);*

Φ^x *is a finite set of* atomic propositions*;*

$\pi^x : Q^x \mapsto 2^{\Phi^x}$ *is an* interpretation function*, which gives the set of primitive propositions satisfied in each state: if $p \in \pi^x(q)$, then this means that the propositional variable p is satisfied (equivalently, true) in state q; and*

$\delta^x : Q^x \times Q^x \times Av^x \mapsto \{+, -, =\}$ *is a* valuation function*, which defines the* status *(promoted ($+$), demoted ($-$), or neutral ($=$)) of a value $v \in Av^x$ ascribed by the agent to the transition between two states: $\delta^x(q, q', v)$ labels the transition between q and q' with respect to the value $v \in Av^x$.*

Note, $Q^x = \emptyset \leftrightarrow Ac^x = \emptyset \leftrightarrow Av^x = \emptyset \leftrightarrow \Phi^x = \emptyset$.

Given its VATS, an agent can now instantiate the practical reasoning argument scheme in order to construct arguments for (or against) actions to achieve a particular goal because they promote (or demote) a particular value.

Definition 2: *An* **argument** *constructed by an agent x from its VATS S^x is a 4-tuple $A = \langle a, p, v, s \rangle$ s.t.: $q_x = q_0^x$; $a \in Ac^x$; $\tau^x(q_x, a) = q_y$; $p \in \pi^x(q_y)$; $v \in Av^x$; $\delta^x(q_x, q_y, v) = s$ where $s \in \{+, -\}$.*
We define the functions: $\mathsf{Act}(A) = a$; $\mathsf{Goal}(A) = p$; $\mathsf{Val}(A) = v$; $\mathsf{Sign}(A) = s$.
If $\mathsf{Sign}(A) = +(-resp.)$, then we say A is an argument **for** *(***against** *resp.) action a.*
We denote the **set of all arguments an agent** *x* **can construct from** *S^x as $Args^x$; we let $Args^x_p = \{A \in Args^x \mid \mathsf{Goal}(A) = p\}$.*
The set of **values** *for a set of arguments $\mathcal{X}$ is defined as $\mathsf{Vals}(\mathcal{X}) = \{v \mid A \in \mathcal{X}$ and $\mathsf{Val}(A) = v\}$.*

If we take a particular argument for an action, it is possible to generate attacks on that argument by posing the various CQs related to the practical reasoning argument scheme. In [8], details are given of how the reasoning with the argument scheme and posing CQs is split into three stages: *problem formulation*, where the agents decide on the facts and values relevant to the particular situation under consideration; *epistemic reasoning*, where the agents determine the current situation with respect to the structure formed at the previous stage; and *action selection*, where the agents develop, and evaluate, arguments and counter arguments about what to do. Here, we assume that the agents' problem formulation and epistemic reasoning are sound and that there is no dispute between them relating to these stages; hence, we do not consider the CQs that arise in these stages. That leaves CQ5-CQ11 for consideration (as numbered in [8]):

CQ5: Are there alternative ways of realising the same consequences?

CQ6: Are there alternative ways of realising the same goal?

CQ7: Are there alternative ways of promoting the same value?

CQ8: Does doing the action have a side effect which demotes the value?

CQ9: Does doing the action have a side effect which demotes some other value?

CQ10: Does doing the action promote some other value?

CQ11: Does doing the action preclude some other action which would promote some other value?

We do not consider CQ5 or CQ11 further, as the focus of the dialogue is to agree to an action that achieves the *goal*; hence, the incidental consequences (CQ5) and other potentially precluded actions (CQ11) are of no interest. We focus instead on CQ6-CQ10; agents participating in a deliberation dialogue use these CQs to identify attacks on proposed arguments for action. These CQs generate a set of arguments for and against different actions to achieve a particular goal, where each argument is associated with a motivating value. To evaluate the status of these arguments we use a Value Based Argumentation Framework (VAF), introduced in [7]. A VAF is an extension of the argumentation frameworks (AF) of Dung [10]. In an AF an argument is admissible with respect to a set of arguments S if all of its attackers are attacked by some argument in S, and no argument in S attacks an argument in S. In a VAF an argument succeeds in defeating an argument it attacks only if its value is ranked as high, or higher, than the value of the argument attacked; a particular ordering of the values is characterised as an *audience*. Arguments in a VAF are admissible with respect to an audience A and a set of arguments S if they are admissible with respect to S in the AF which results from removing all the attacks which are unsuccessful given the audience A. A maximal admissible set of a VAF is known as a *preferred extension*.

Although VAFs are commonly defined abstractly, here we give an instantiation in which we define the attack relation between the arguments. Condition 1 of the following attack relation allows for CQ8 and CQ9; condition 2 allows for CQ10; condition 3 allows for CQ6 and CQ7. Note that attacks generated by condition 1 are not symmetrical, whilst those generated by conditions 2 and 3 are.

Definition 3: *An* **instantiated value-based argumentation framework (iVAF)** *is defined by a tuple* $\langle \mathcal{X}, \mathcal{A} \rangle$ *s.t.* $\mathcal{X}$ *is a finite set of arguments and* $\mathcal{A} \subset \mathcal{X} \times \mathcal{X}$ *is the* **attack relation**. *A pair* $(A_i, A_j) \in \mathcal{A}$ *is referred to as "*A_i *attacks* A_j*" or "*A_j *is attacked by*

A_i". *For two arguments* $A_i = \langle a, p, v, s\rangle$, $A_j = \langle a', p', v', s'\rangle \in \mathcal{X}$, $(A_i, A_j) \in \mathcal{A}$ *iff* $p = p'$ *and either:*

1. $a = a'$, $s = -$ *and* $s' = +$; *or*
2. $a = a'$, $v \neq v'$ *and* $s = s' = +$; *or*
3. $a \neq a'$ *and* $s = s' = +$.

An **audience** *for an agent* x *over the values* V *is a binary relation* $\mathcal{R}^x \subset V \times V$ *that defines a* total order *over* V. *We say that an argument* A_i *is* **preferred to** *the argument* A_j *in the audience* $\mathcal{R}^x$, *denoted* $A_i \succ_x A_j$, *iff* $(\mathsf{Val}(A_i), (\mathsf{Val}(A_j)) \in \mathcal{R}^x$. *If* R^x *is an audience over the values* V *for the iVAF* $\langle \mathcal{X}, \mathcal{A}\rangle$, *then* $\mathsf{Vals}(\mathcal{X}) \subseteq V$.

We use the term audience here to be consistent with the literature, it does not refer to the preference of a *set* of agents; rather, we define it to represent a particular agent's preference over a set of values.

Given an iVAF and a particular agent's audience, we can determine acceptability of an argument as follows. Note that if an attack is symmetric, then an attack only succeeds in defeat if the attacker is more preferred than the argument being attacked; however, as in [7], if an attack is asymmetric, then an attack succeeds in defeat if the attacker is at least as preferred as the argument being attacked.

Definition 4: *Let* $\mathcal{R}^x$ *be an audience and let* $\langle \mathcal{X}, \mathcal{A}\rangle$ *be an iVAF.*

For $(A_i, A_j) \in \mathcal{A}$ *s.t.* $(A_j, A_i) \notin \mathcal{A}$, A_i **defeats** A_j *under* $\mathcal{R}^x$ *if* $A_j \not\succ_x A_i$.

For $(A_i, A_j) \in \mathcal{A}$ *s.t.* $(A_j, A_i) \in \mathcal{A}$, A_i **defeats** A_j *under* $\mathcal{R}^x$ *if* $A_i \succ_x A_j$.

An argument $A_i \in \mathcal{X}$ *is* **acceptable w.r.t** S *under* $\mathcal{R}^x$ *(*$S \subseteq \mathcal{X}$*) if: for every* $A_j \in \mathcal{X}$ *that defeats* A_i *under* $\mathcal{R}^x$, *there is some* $A_k \in S$ *that defeats* A_j *under* $\mathcal{R}^x$.

A subset S *of* $\mathcal{X}$ *is* **conflict-free** *under* $\mathcal{R}^x$ *if no argument* $A_i \in S$ *defeats another argument* $A_j \in S$ *under* $\mathcal{R}^x$.

A subset S *of* $\mathcal{X}$ *is* **admissible** *under* $\mathcal{R}^x$ *if:* S *is conflict-free in* $\mathcal{R}^x$ *and every* $A \in S$ *is acceptable w.r.t* S *under* $\mathcal{R}^x$.

A subset S *of* $\mathcal{X}$ *is a* **preferred extension** *under* $\mathcal{R}^x$ *if it is a maximal admissible set under* $\mathcal{R}^x$.

An argument A *is* **acceptable** *in the iVAF* $\langle \mathcal{X}, \mathcal{A}\rangle$ *under audience* $\mathcal{R}^x$ *if there is* some *preferred extension containing it.*

We have now defined a mechanism with which an agent can determine attacks between arguments for and against actions, and can then use an ordering over the values that motivate such arguments (its audience) in order to determine their acceptability. In the next section we define our dialogue system.

3 Dialogue system

The communicative acts in a dialogue are called *moves*. We assume that there are always exactly two agents (*participants*) taking part in a dialogue, each with its own identifier taken from the set $\mathcal{I} = \{1, 2\}$. Each participant takes it in turn to make a move to the other participant. We refer to participants using the variables x and $\overline{x}$ such that: x is 1 if and only if $\overline{x}$ is 2; x is 2 if and only if $\overline{x}$ is 1.

Move	Format
$open$	$\langle x, \mathsf{open}, \gamma \rangle$
$assert$	$\langle x, \mathsf{assert}, A \rangle$
$agree$	$\langle x, \mathsf{agree}, a \rangle$
$close$	$\langle x, \mathsf{close}, \gamma \rangle$

Table 1. Format for moves used in deliberation dialogues: γ is a goal; a is an action; A is an argument; $x \in \{1, 2\}$ is an agent identifier.

A move in our system is of the form $\langle Agent, Act, Content \rangle$. $Agent$ is the identifier of the agent generating the move, Act is the type of move, and the $Content$ gives the details of the move. The format for moves used in deliberation dialogues is shown in Table 1, and the set of all moves meeting the format defined in Table 1 is denoted $\mathcal{M}$. Note that the system allows for other types of dialogues to be generated and these might require the addition of extra moves. Also, $\mathsf{Sender} : \mathcal{M} \mapsto \mathcal{I}$ is a function such that $\mathsf{Sender}(\langle Agent, Act, Content \rangle) = Agent$.

We now informally explain the different types of move: an *open* move $\langle x, \mathsf{open}, \gamma \rangle$ opens a dialogue to agree on an action to achieve the goal γ; an *assert* move $\langle x, \mathsf{assert}, A \rangle$ asserts an argument A for or against an action to achieve a goal that is the topic of the dialogue; an *agree* move $\langle x, \mathsf{agree}, a \rangle$ indicates that x agrees to performing action a to achieve the topic; a *close* move $\langle x, \mathsf{close}, \gamma \rangle$ indicates that x wishes to end the dialogue.

A dialogue is simply a sequence of moves, each of which is made from one participant to the other. As a dialogue progresses over time, we denote each timepoint by a natural number. Each move is indexed by the timepoint when the move was made. Exactly one move is made at each timepoint.

Definition 5: *A* **dialogue**, *denoted D^t, is a sequence of moves $[m_1, \ldots, m_t]$ involving two participants in $\mathcal{I} = \{1, 2\}$, where $t \in \mathbb{N}$ and the following conditions hold:*

1. m_1 *is a move of the form* $\langle x, \mathsf{open}, \gamma \rangle$ *where* $x \in \mathcal{I}$
2. $\mathsf{Sender}(m_s) \in \mathcal{I}$ *for* $1 \leq s \leq t$
3. $\mathsf{Sender}(m_s) \neq \mathsf{Sender}(m_{s+1})$ *for* $1 \leq s < t$

The **topic** *of the dialogue D^t is returned by* $\mathsf{Topic}(D^t) = \gamma$. *The set of all dialogues is denoted $\mathcal{D}$.*

The first move of a dialogue D^t must always be an open move (condition 1 of the previous definition), every move of the dialogue must be made by a participant (condition 2), and the agents take it in turns to send moves (condition 3). In order to terminate a dialogue, either: two close moves must appear one immediately after the other in the sequence (a *matched-close*); or two moves agreeing to the same action must appear one immediately after the other in the sequence (an *agreed-close*).

Definition 6: *Let D^t be a dialogue s.t.* $\mathsf{Topic}(D^t) = \gamma$. *We say that m_s ($1 < s \leq t$), is*

- *a* **matched-close for** D^t *iff* $m_{s-1} = \langle x, \mathsf{close}, \gamma \rangle$ *and* $m_s = \langle \overline{x}, \mathsf{close}, \gamma \rangle$.
- *an* **agreed-close for** D^t *iff* $m_{s-1} = \langle x, \mathsf{agree}, a \rangle$ *and* $m_s = \langle \overline{x}, \mathsf{agree}, a \rangle$.

We say D^t has a **failed outcome** *iff m_t is a matched-close, whereas we say D^t has a* **successful outcome** *of a iff $m_t = \langle x, \mathsf{agree}, a \rangle$ is an agreed-close.*

So a matched-close or an agreed-close will terminate a dialogue D^t but only if D^t has not already terminated.

Definition 7: *Let D^t be a dialogue. D^t* **terminates at** *t iff m_t is a matched-close or an agreed-close for D^t and $\neg\exists s$ s.t. $s < t$, D^t* **extends** *D^s (i.e. the first s moves of D^t are the same as the sequence D^s) and D^s terminates at s.*

We shortly give the particular protocol and strategy functions that allow agents to generate deliberation dialogues. First, we introduce some subsidiary definitions. At any point in a dialogue, an agent x can construct an iVAF from the union of the arguments it can construct from its VATS and the arguments that have been asserted by the other agent; we call this x's *dialogue iVAF*.

Definition 8: *A* **dialogue iVAF** *for an agent x participating in a dialogue D^t is denoted* $\text{dVAF}(x, D^t)$. *If D^t is the sequence of moves $= [m_1, \ldots, m_t]$, then* $\text{dVAF}(x, D^t)$ *is the iVAF $\langle \mathcal{X}, \mathcal{A} \rangle$ where $\mathcal{X} = Args^x_{\mathsf{Topic}(D^t)} \cup \{A \mid \exists m_k = \langle \overline{x}, assert, A \rangle (1 \leq k \leq t)\}$.*

An action is *agreeable* to an agent x if and only if there is some argument *for* that action that is acceptable in x's dialogue iVAF under the audience that represents x's preference over values. Note that the set of actions that are agreeable to an agent may change over the course of the dialogue.

Definition 9: *An action a is* **agreeable** *in the iVAF $\langle \mathcal{X}, \mathcal{A} \rangle$ under the audience $\mathcal{R}^x$ iff $\exists A = \langle a, \gamma, v, + \rangle \in \mathcal{X}$ s.t. A is acceptable in $\langle \mathcal{X}, \mathcal{A} \rangle$ under $\mathcal{R}^x$. We denote the* **set of all actions that are agreeable to an agent** *x* **participating in a dialogue** *D^t as* $\mathsf{AgActs}(x, D^t)$, *s.t. $a \in \mathsf{AgActs}(x, D^t)$ iff a is agreeable in* $\text{dVAF}(x, D^t)$ *under $\mathcal{R}^x$.*

A protocol is a function that returns the set of moves that are permissible for an agent to make at each point in a particular type of dialogue. Here we give a deliberation protocol. It takes the dialogue that the agents are participating in and the identifier of the agent whose turn it is to move, and returns the set of permissible moves.

Definition 10: *The* **deliberation protocol** *for agent x is a function $\mathsf{Protocol}_x : \mathcal{D} \mapsto \wp(\mathcal{M})$. Let D^t be a dialogue ($1 \leq t$) with participants $\{1, 2\}$ s.t. $\mathsf{Sender}(m_t) = \overline{x}$ and $\mathsf{Topic}(D^t) = \gamma$.*

$$\mathsf{Protocol}_x(D^t) = P^{\mathsf{ass}}_x(D^t) \cup P^{\mathsf{ag}}_x(D^t) \cup \{\langle x, \mathsf{close}, \gamma \rangle\}$$

where the following are sets of moves and $x' \in \{1, 2\}$.

$$\begin{aligned} P^{\mathsf{ass}}_x(D^t) = \{\langle x, \mathsf{assert}, A \rangle \mid\ & \mathsf{Goal}(A) = \gamma \\ & \textbf{and} \\ & \neg\exists m_{t'} = \langle x', assert, A \rangle (1 < t' \leq t) \end{aligned}$$

$$\begin{aligned} P^{\mathsf{ag}}_x(D^t) = \{\langle x, \mathsf{agree}, a \rangle \mid\ & \textbf{either} \\ & (1) m_t = \langle \overline{x}, agree, a \rangle\} \\ & \textbf{else} \\ & (2)(\exists m_{t'} = \langle \overline{x}, assert, \langle a, \gamma, v, + \rangle \rangle (1 < t' \leq t) \\ & \quad \textbf{and} \\ & \quad (\ \textbf{if } \exists m_{t''} = \langle x, agree, a \rangle) \\ & \quad \textbf{then } \exists A, m_{t'''} = \langle x, assert, A \rangle \\ & \quad\ (t'' < t''' \leq t)))\} \end{aligned}$$

The protocol states that it is permissible to assert an argument as long as that argument has not previously been asserted in the dialogue. An agent can agree to an action

$$\mathsf{Strategy}_x(D^t) = \begin{cases} \mathsf{Pick}(S_x^{\mathsf{ag}})(D^t) & \text{iff } S_x^{\mathsf{ag}}(D^t) \neq \emptyset \\ \mathsf{Pick}(S_x^{\mathsf{prop}})(D^t) & \text{iff } S_x^{\mathsf{ag}}(D^t) = \emptyset \text{ and } S_x^{\mathsf{prop}}(D^t) \neq \emptyset \\ \mathsf{Pick}(S_x^{\mathsf{att}})(D^t) & \text{iff } S_x^{\mathsf{ag}}(D^t) = S_x^{\mathsf{prop}}(D^t) = \emptyset \text{ and } S_x^{\mathsf{att}}(D^t) \neq \emptyset \\ \langle x, \mathsf{close}, \mathsf{Topic}(D^t)\rangle & \text{iff } S_x^{\mathsf{ag}}(D^t) = S_x^{\mathsf{prop}}(D^t) = S_x^{\mathsf{att}}(D^t) = \emptyset \end{cases}$$

where the choices for the moves are given by the following subsidiary functions ($x' \in \{x, \overline{x}\}, \mathsf{Topic}(D^t) = \gamma$):

$$\begin{aligned} S_x^{\mathsf{ag}}(D^t) &= \{\langle x, \mathsf{agree}, a\rangle \in P_x^{\mathsf{ag}}(D^t) \mid a \in \mathsf{AgActs}(x, D^t)\} \\ S_x^{\mathsf{prop}}(D^t) &= \{\langle x, \mathsf{assert}, A\rangle \in P_x^{\mathsf{ass}}(D^t) \mid A \in Args_\gamma^x, \mathsf{Act}(A) = a, \mathsf{Sign}(A) = + \text{ and} \\ &\qquad a \in \mathsf{AgActs}(x, D^t)\} \\ S_x^{\mathsf{att}}(D^t) &= \{\langle x, \mathsf{assert}, A\rangle \in P_x^{\mathsf{ass}}(D^t) \mid A \in Args_\gamma^x, \mathsf{Act}(A) = a, \mathsf{Sign}(A) = -, \\ &\qquad a \notin \mathsf{AgActs}(x, D^t) \text{ and } \exists m_{t'} = \langle x', \mathsf{assert}, A'\rangle \\ &\qquad (1 \leq t' \leq t) \text{ s.t. } \mathsf{Act}(A') = a \text{ and } \mathsf{Sign}(A') = +\} \end{aligned}$$

Fig. 1. The **strategy** function uniquely selects a move according to the following preference ordering (starting with the most preferred): an agree move (ag), a proposing assert move (prop), an attacking assert move (att), a close move (close).

that has been agreed to by the other agent in the preceding move (condition 1 of P_x^{ag}); otherwise an agent x can agree to an action that has been proposed by the other participant (condition 2 of P_x^{ag}) as long as if x has previously agreed to that action, then x has since then asserted some new argument. This is because we want to avoid the situation where an agent keeps repeatedly agreeing to an action that the other agent will not agree to: if an agent makes a move agreeing to an action and the other agent does not wish to also agree to that action, then the first agent must introduce some new argument that may convince the second agent to agree before being able to repeat its agree move. Agents may always make a close move. Note, it is straightforward to check conformance with the protocol as it only refers to public elements of the dialogue.

We now define a *basic deliberation strategy*. It takes the dialogue D^t and returns exactly one of the permissible moves. Note, this strategy makes use of a function $\mathsf{Pick} : \wp(\mathcal{M}) \mapsto \mathcal{M}$. We do not define Pick here but leave it as a parameter of our strategy (in its simplest form Pick may return an arbitrary move from the input set); hence our system could generate more than one dialogue depending on the definition of the Pick function. In future work, we plan to design particular Pick functions; for example, taking into account an agent's perception of the other participant (more in section 5).

Definition 11: *The* **basic strategy** *for an agent x is a function* $\mathsf{Strategy}_x : \mathcal{D} \mapsto \mathcal{M}$ *given in Figure 1.*

A *well-formed deliberation dialogue* is a dialogue that has been generated by two agents each following the basic strategy.

Definition 12: *A* **well-formed deliberation dialogue** *is a dialogue D^t s.t. $\forall t'$ ($1 \leq t' \leq t$),* $\mathsf{Sender}(m^{t'}) = x$ *iff* $\mathsf{Strategy}_x(D^{t'-1}) = m_{t'}$

We now present a simple example. There are two participating agents ($\{1, 2\}$) who have the joint goal to go out for dinner together (din). $Ac^1 \cup Ac^2 = \{it, ch\}$ (it: go to an Italian restaurant; ch: go to a Chinese restaurant) and $Av^1 \cup Av^2 = \{d, e1, e2, c\}$ (d: distance to travel; $e1$: agent 1's enjoyment; $e2$: agent 2's enjoyment; c: cost). The agents' audiences are as follows.

$$d \succ_1 e1 \succ_1 c \succ_1 e2$$
$$c \succ_2 e2 \succ_2 e1 \succ_2 d$$

Agent 1 starts the dialogue.

$$m_1 = \langle 1, open, din \rangle$$

The agents' dialogue iVAFs at this opening stage in the dialogue can be seen in Figs. 2 and 3, where the nodes represent arguments and are labelled with the action that they are for (or the negation of the action that they are against) and the value that they are motivated by. The arcs represent the attack relation between arguments, and a double circle round a node means that the argument it represents is acceptable to that agent.

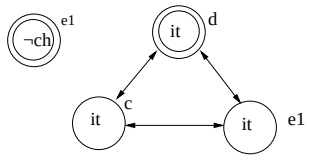

Fig. 2. Agent 1's dialogue iVAF at t = 1, *dVAF*$(1, D^1)$.

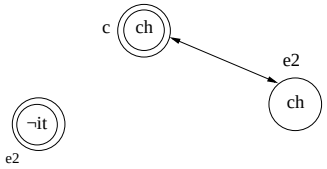

Fig. 3. Agent 2's dialogue iVAF at t = 1, *dVAF*$(2, D^1)$.

At this point in the dialogue, there is only one argument *for* an action that is acceptable to 2 ($\langle ch, din, c, + \rangle$), hence ch is the only action that is agreeable to 2. 2 must therefore assert an argument that it can construct for going to the Chinese restaurant. There are two such arguments that the Pick function could select ($\langle ch, din, c, + \rangle$, $\langle ch, din, e2, + \rangle$). Let us assume that $\langle ch, din, c, + \rangle$ is selected.

$$m_2 = \langle 2, assert, \langle ch, din, c, + \rangle \rangle$$

This new argument is added to 1's dialogue iVAF, to give *dVAF*$(1, D^2)$ (Fig. 4).

Although agent 2 has proposed going to the Chinese restaurant, this action is not agreeable to agent 1 at this point in the dialogue (as there is no argument for this action that is acceptable in Fig. 4). There is, however, an argument for the action it ($\langle it, din, d, + \rangle$) that is acceptable in 1's dialogue iVAF (Fig. 4), and so going to the Italian restaurant is agreeable to 1. Hence, 1 must make an assert move proposing an argument for the action it, and there are three such arguments that the Pick function can select from ($\langle it, din, d, + \rangle$, $\langle it, din, c, + \rangle$, $\langle it, din, e1, + \rangle$). Let us assume that $\langle it, din, c, + \rangle$ is selected.

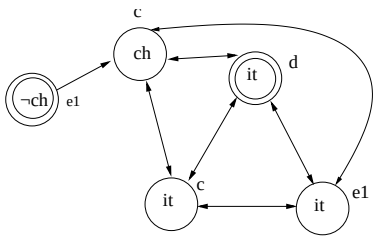

Fig. 4. Agent 1's dialogue iVAF at t = 2, *dVAF*$(1, D^2)$.

$$m_3 = \langle 1, assert, \langle it, din, c, + \rangle \rangle$$

This new argument is added to 2's dialogue iVAF, to give *dVAF*$(2, D^3)$ (Fig. 5).

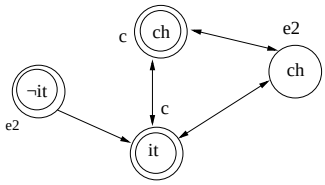

Fig. 5. Agent 2's dialogue iVAF at t = 3, *dVAF*$(2, D^3)$.

Going to the Italian restaurant is now agreeable to agent 2 since the new argument introduced promotes the value ranked most highly for agent 2, i.e. cost, and so this argument is acceptable. So, 2 agrees to this action.

$$m_4 = \langle 2, agree, it \rangle$$

Going to the Italian restaurant is also agreeable to agent 1 (as the argument $\langle it, din, d, + \rangle$ is acceptable in its dialogue iVAF, which is still the same as that shown in Fig. 4 as 2 has not asserted any new arguments), hence 1 also agrees to this action.

$$m_5 = \langle 1, agree, it \rangle$$

Note that the dialogue has terminated successfully and the agents are each happy to agree to go to the Italian restaurant; however, this action is agreeable to each agent for a different reason. Agent 1 is happy to go to the Italian restaurant as it promotes the value of distance to travel (the Italian restaurant is close by), whereas agent 2 is happy to go to the Italian restaurant as it will promote the value of cost (as it is a cheap restaurant). The agents need not be aware of one another's audience in order to reach an agreement.

It is worth mentioning that, as we have left the Pick function unspecified, our strategy could have generated a longer dialogue if, for example, agent 1 had instead chosen to assert the argument $\langle it, din, d, + \rangle$ at the move m_3. This illustrates how an agent's perception of the other participant may be useful: in the previous example agent 1 may make the assumption that, as agent 2 has previously asserted an argument that promotes

cost, cost is something that agent 2 values; or an agent may use its perception of another agent's personality to guide argument selection [11].

Another point to note concerns the arguments generated by CQ10. Such arguments do not dispute that the action should be performed, but do dispute the reasons as to why, and so they are modelled as attacks despite being for the same action. Pinpointing this distinction here is important for two main reasons. Firstly, an advantage of the argumentation approach is that agents make explicit the reasons as to why they agree and disagree about the acceptability of arguments, and the acceptability may well turn on such reasons. Where there are two arguments proposed for the same action but each is based upon different values, an agent may only accept the argument based on one of the values. Hence such arguments are seen to be in conflict. Secondly, by participating in dialogues agents reveal what their value orderings are, as pointed out in [12]. If an agent will accept an argument for action based upon one particular value but not another, then this is potentially useful information for future dialogue interactions; if agreement is not reached about a particular action proposal, then dialogue participants will know the values an opposing agent cares about and this can guide the selection of further actions to propose, as we discuss later on in section 5.

A final related issue to note is that of accrual of arguments. If there are multiple arguments for an action and the values promoted are acceptable to the agents then some form of accrual might seem desirable. However, the complex issue of how best to accrue such arguments has not been fully resolved and this is not the focus here.

4 Properties

Certainly (assuming the cooperative agents do not abandon the dialogue for some reason), all dialogues generated by our system terminate. This is clear as we assume that the sets of actions and values available to an agent are finite, hence the set of arguments that an agent can construct is also finite. As the protocol does not allow the agents to keep asserting the same argument, or to keep agreeing to the same action unless a new argument has been asserted, either the dialogue will terminate successfully else the agents will run out of legal assert and agree moves and so each will make a close move.

Proposition 1: *If D^t is a well-formed deliberation dialogue, then $\exists t'$ ($t \leq t'$) s.t. $D^{t'}$ is a well-formed deliberation dialogue that terminates at t' and $D^{t'}$ extends D^t.*

It is also the clear from the definition of the strategy (which only allows an action to be agreed to if that action is agreeable to the agent) that if the dialogue terminates with a successful outcome of action a, then a is agreeable to both agents.

Proposition 2: *If D^t is a well-formed deliberation dialogue that terminates successfully at t with outcome a, then $a \in AgActs(x, D^t)$ and $a \in AgActs(\overline{x}, D^t)$.*

Similarly, we can show that if there is an action that is agreeable to both agents when the dialogue terminates, then the dialogue will terminate successfully. In order to show this, however, we need a subsidiary lemma that states: if an agent makes a close move, then any arguments that it can construct that are for actions that it finds agreeable must have been asserted by one of the agents during the dialogue. This follows from the definition of the strategy, which only allows agents to make a close move once they have exhausted all possible assert moves.

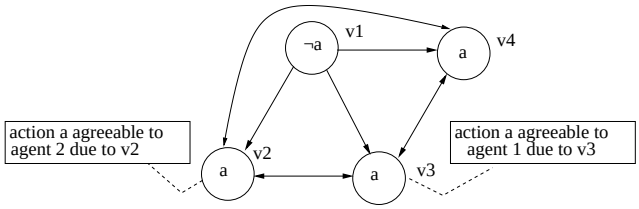

Fig. 6. The joint iVAF

Lemma 1: *Let D^t be a well-formed deliberation dialogue with* $\mathsf{Topic}(D^t) = \gamma$*, s.t.* $m_t = \langle x, close, \gamma \rangle$ *and* dVAF$(x, D^t) = \langle \mathcal{X}, \mathcal{A} \rangle$*. If* $A = \langle a, \gamma, v, + \rangle \in \mathcal{X}$ *and* $a \in AgActs(x, D^t)$*, then* $\exists m_{t'} = \langle x', assert, A, \rangle$ *(*$1 < t' \leq t$*,* $x' \in \{x, \overline{x}\}$*).*

Now we show that if there is an action that is agreeable to both agents when the dialogue terminates, then the dialogue will have a successful outcome.

Proposition 3: *Let D^t be a well-formed deliberation dialogue that terminates at t. If $a \in AgActs(x, D^t)$ and $a \in AgActs(\overline{x}, D^t)$, then D^t terminates successfully.*

Proof: *Assume that D^t terminates unsuccessfully at t and that* $\mathsf{Sender}(m_t) = \overline{x}$*. From Lemma 1, there is at least one argument A for a that has been asserted by one of the agents. There are two cases. Case 1: x asserted A. Case 2: $\overline{x}$ asserted A.*

Case 1: x asserted A. Hence (from the protocol) it would have been legal for $\overline{x}$ to make the move $m_t = \langle \overline{x}, agree, a \rangle$ (in which case x would have had to replied with an agree, giving successful termination), unless $\overline{x}$ had previously made a move $m_{t'} = \langle \overline{x}, agree, a \rangle$ but had not made a move $m_{t''} = \langle \overline{x}, assert, A \rangle$ with $t' < t'' < t$. However, if this were the case, then we would have $AgActs(x, D^{t'}) = AgActs(x, D^t)$ (because no new arguments have been put forward by $\overline{x}$ to change x's dialogue iVAF), hence x would have had to respond to the move $m_{t'}$ with an agree, terminating the dialogue successfully. Hence contradiction.

Case 2: works equivalently to case 1. Hence, D^t terminates successfully. □

We have shown then: all dialogues terminate; if a dialogue terminates successfully, then the outcome will be agreeable to both participants; if a dialogue terminates and there is some action that is agreeable to both agents, then the dialogue will have a successful outcome.

It would be desirable to show that if there is some action that is agreeable in the **joint iVAF**, which is the iVAF that can be constructed from the union of the agents' arguments (i.e. the iVAF $\langle \mathcal{X}, \mathcal{A} \rangle$, where $\mathcal{X} = Args^x_\gamma \cup Args^{\overline{x}}_\gamma$ and γ is the topic of the dialogue), then the dialogue will terminate successfully. However, there are some cases where there is an action that is agreeable in the joint iVAF to each of the participants and yet still they may not reach an agreement. Consider the following example in which there is an action a that is agreeable to both the agents given the joint iVAF (see Fig.6) and yet the dialogue generated here terminates unsuccessfully.

The participants ($\{1, 2\}$) have the following audiences.

$$v3 \succ_1 v1 \succ_1 v4 \succ_1 v2$$
$$v2 \succ_2 v1 \succ_2 v4 \succ_2 v3$$

Agent 1 starts the dialogue.

$$m_1 = \langle 1, open, p \rangle$$

The agents' dialogue iVAFs at this stage in the dialogue can be seen in Figs. 7 and 8.

Fig. 7. Agent 1's dialogue iVAF at t = 1, *dVAF*$(1, D^1)$.

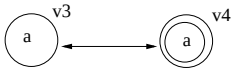

Fig. 8. Agent 2's dialogue iVAF at t = 1, *dVAF*$(2, D^1)$.

At this point in the dialogue, there is one action that is agreeable to agent 2 (a, as there is an argument *for* a that is acceptable in Fig. 8); hence (following the basic dialogue strategy), agent 2 must assert one of the arguments that it can construct for a (either $\langle a, p, v3, + \rangle$ or $\langle a, p, v4, + \rangle$). Recall, we have not specified the Pick function that has to choose between these two possible proposing assert moves. Let us assume that the Pick function makes an arbitrary choice to assert $\langle a, p, v4, + \rangle$.

$$m_2 = \langle 2, assert, \langle a, p, v4, + \rangle \rangle$$

This new argument is added to agent 1's dialogue iVAF, to give *dVAF*$(1, D^2)$ (Fig. 9).

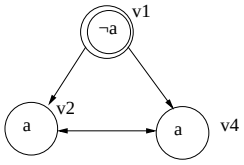

Fig. 9. Agent 1's dialogue iVAF at t = 2, *dVAF*$(1, D^2)$.

From Fig. 9, we see that the only argument that is now acceptable to agent 1 is the argument *against* a ($\langle a, p, v1, - \rangle$), hence there are no actions that are agreeable to agent 1. Thus agent 1 must make an attacking assert move.

$$m_3 = \langle 1, assert, \langle a, p, v1, - \rangle \rangle$$

This new argument is added to agent 2's dialogue iVAF, to give *dVAF*$(2, D^3)$ (Fig. 10).

We see from Fig. 10 that the only argument that is now acceptable to agent 2 is the argument *against* a that 1 has just asserted ($\langle a, p, v1, - \rangle$); hence, a is now no longer an agreeable action for agent 2. As there are now no actions that are agreeable to agent

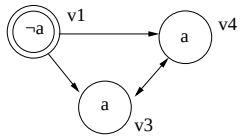

Fig. 10. Agent 2's dialogue iVAF at t = 3, *dVAF*$(2, D^3)$.

2, it cannot make any proposing assert moves. It also cannot make any attacking assert moves, as the only argument that it can construct against an action has already been asserted by agent 1. Hence, agent 2 makes a close move.

$$m_4 = \langle 2, close, p \rangle$$

Thus, the dialogue iVAF for 1 is still the same as that which appears in Fig. 9. As there are no actions that are agreeable to agent 1, it cannot make any proposing assert moves. It cannot make any attacking assert moves, as the only argument that it can construct against an action has already been asserted. Hence, agent 1 also makes a close move.

$$m_5 = \langle 1, close, p \rangle$$

The dialogue has thus terminated unsuccessfully and the agents have not managed to reach an agreement as to how to achieve the goal p. However, we can see that if the Pick function instead selected the argument $\langle a, p, v3, + \rangle$ for agent 2 to assert for the move m_2, then the resulting dialogue would have led to a successful outcome.

This example then illustrates a particular problem: the arguments exist that will enable the agents to reach an agreement (we can see this in the joint iVAF, Fig. 6, in which each agent finds a agreeable) and yet the particular arguments selected by the Pick function may not allow agreement to be reached. The choice of moves made in a deliberation dialogue affects the dialogue outcome; hence, strategic manoeuvring within the dialogue is possible in order to try to influence the dialogue outcome.

This evaluation helps us to understand the complex issues and difficulties involved in allowing agents with different preferences to agree how to act. We discuss possible responses to some of these difficulties in the next section.

5 Proposed extensions

One way in which we could aim to avoid the problem illustrated in the previous example is by allowing agents to develop a model of which values they believe are important to the other participant. This model can then be used by the Pick function in order to select arguments that are more likely to lead to agreement (i.e. those that the agent believes promote or demote values that are highly preferred by the other participant). Consider the above example, if agent 2 believed that value $v3$ was more preferred to agent 1 than value $v4$, then 2 would have instead asserted $\langle a, p, v3, + \rangle$ for the move m_2, which would have led to a successful outcome.

Therefore, the first extension that we plan to investigate is to design a particular Pick function that takes into account what values the agent believes are important to the other

participant. We also plan to develop a mechanism which allows the agent to build up its model of the other participant, based on the other participant's dialogue behaviour; for example, if an agent x asserts an argument for an action a because it promotes a particular value v, and the other participant $\overline{x}$ does not then agree to a, agent x may have reason to believe that $\overline{x}$ does not highly rank the value v.

Another problem that may be faced with our dialogue system is when it is not possible for the agents to come to an agreement no matter which arguments they choose to assert. The simplest example of this is when each agent can only construct one argument to achieve the topic p: agent 1 can construct $\langle a1, p, v1, + \rangle$; agent 2 can construct $\langle a2, p, v2, + \rangle$. Now if agent 1's audience is such that it prefers $v1$ to $v2$ and agent 2's audience is such that it prefers $v2$ to $v1$, then the agents will not be able to reach an agreement with the dialogue system that we have proposed here; this is despite the fact that both agents do share the goal of coming to some agreement on how to act to achieve p. The agents in this case have reached an impasse, where there is no way of finding an action that is agreeable to both agents given their individual preferences over the values.

The second extension that we propose to investigate aims to overcome such an impasse when agreement is nevertheless necessary. We plan to define a new type of dialogue (which could be embedded within the deliberation dialogue we have defined here) that allows the agents to discuss their preferences over the values and to suggest and agree to compromises that allow them to arrive at an agreement in the deliberation dialogue. For example, if agent 1's audience is $v1 \succ_1 v2 \succ_1 v3$ and agent 2's audience is $v3 \succ_2 v2 \succ_2 v1$, then they may both be willing to switch their first and second most preferred values if this were to lead to an agreement (i.e. giving $v2 \succ_1 v1 \succ_1 v3$ and $v2 \succ_2 v3 \succ_2 v1$).

We would also like to extend our system to deal with the situation in which the other stages of practical reasoning (problem formulation and epistemic reasoning) may be flawed. In [5], an approach to dealing with epistemic reasoning was presented, that allowed an embedded inquiry subdialogue with which agents could jointly reason epistemically about the state of the world. Thus, the third extension that we propose is to develop a new type of dialogue that will allow agents to jointly reason about the elements of a VATS in order to consider possible flaws in the problem formulation stage.

6 Related Work

There is existing work in the literature on argumentation that bears some relation to what we have presented here, though the aims and contributions of these approaches are markedly different.

Our proposal follows the approach in [5, 13] but the types of moves are different, and the protocol and strategy functions are substantially altered from those presented in either [5] or [13]. This alteration is necessary as neither of [5, 13] allow agents to participate in deliberation dialogues. In [13], a dialogue system is presented for epistemic inquiry dialogues; it allows agents to jointly construct argument graphs (where the arguments refer only to beliefs) and to use a shared defeat relation to determine the acceptability of particular arguments.

The proposal of [5] is closer to that presented here, as both are concerned with how to act. However, the dialogue system in [5] does not allow deliberation dialogues as the outcome of any dialogue that it generates is predetermined by the union of the participating agents' knowledge. Rather, the dialogues of [5] are better categorised as a joint inference; they ensure that the agents assert all arguments that may be relevant to the question of how to act, after which a universal value ordering is applied to determine the outcome. As a shared universal value ordering is used in [5], there is an objective view of the "best" outcome (being that which you would get if you pooled the agents' knowledge and applied the shared ordering); this is in contrast to the dialogue system we present here, where the "best" outcome is subjective and depends on the point of view of a particular agent. As the agents presented here each have their own distinct audience, they must come to an explicit agreement about how to act (hence the introduction of an agree move) despite the fact that their internal views of argument acceptability may conflict. Also, here we define the attack relation (in the iVAF), which takes account of the relevant CQs, whilst in [5] the attack relation is only informally discussed.

Deliberation dialogues have only been considered in detail by the authors of [2, 3]. Unlike in our work, in [2] the evaluation of arguments is not done in terms of argumentation frameworks, and strategies for reaching agreement are not considered; and in [3] the focus is on goal selection and planning.

In [12] issues concerning audiences in argumentation frameworks are addressed where the concern is to find particular audiences (if they exist) for which some arguments are acceptable and others are not. Also considered is how preferences over values emerge through a dialogue; this is demonstrated by considering how two agents can make moves within a dialogue where both are dealing with the same joint graph. However, the graph can be seen as a static structure within which agents are playing moves, i.e. putting forward acceptable arguments, rather than constructing a graph that is not complete at the outset, as in the approach we have presented.

There is also some work that considers how Dungian argumentation frameworks associated with individual agents can be merged together [14]. The merging is done not through taking the union of the individual frameworks, but through the application of criteria that determine when arguments and attacks between them can be merged into a larger graph. The main goal of the work is to characterise the sets of arguments acceptable by the whole group of agents using notions of joint acceptability, which include voting methods. In our work we are not interested in merging individual agent's graphs *per se*; rather, an agent develops its own individual graph and uses this to determine if it finds an action agreeable. In [14] no dialogical interactions are considered, and it is also explicitly noted that consideration has not been given to how the merging approach can be applied to value-based argument systems.

Prakken [15] considers how agents can come to a public agreement despite their internal views of argument acceptability conflicting, allowing them to make explicit attack and surrender moves. However, Prakken does not explicitly consider value-based arguments, nor does he discuss particular strategies.

Strategic argumentation has been considered in other work. For example, in [16] a dialogue game for persuasion is presented that is based upon one originally proposed in [1] but makes use of Dungian argumentation frameworks. Scope is provided for

three strategic considerations which concern: revealing inconsistencies between an opponent's commitments and his beliefs; exploiting the opponent's reasoning so as to create such inconsistencies; and revealing blunders to be avoided in expanding the opponent's knowledge base. These strategies all concern reasoning about an opponent's beliefs, as opposed to reasoning about action proposals with subjective preferences, as done in our work, and the game in [16] is of an adversarial nature, whereas our setting is more co-operative.

One account that does consider strategies when reasoning with value-based arguments is given in [7], where the objective is to create obligations on the opponent to accept some argument based on his previously expressed preferences. The starting point for such an interaction is a fixed joint VAF, shared by the dialogue participants. In our approach the information is not centralised in this manner, the argument graphs are built up as the dialogue proceeds, we do not assume perfect knowledge of the other agent's graph and preferences, and our dialogues have a more co-operative nature.

A related new area that is starting to receive attention is the application of game theory to argumentation (e.g. [17]). This work has investigated situations under which rational agents will not have any incentive to lie about or hide arguments; although this is concerned mainly with protocol design, it appears likely that such work will have implications for strategy design.

A few works do explicitly consider the selection of dialogue targets, that is the selection of a particular previous move to respond to. In [15] a move is defined as relevant if its target would (if attacked) cause the status of the original move to change; properties of dialogues are considered where agents are restricted to making relevant moves. In [18] this is built on to consider other classes of move relevance and the space that agents then have for strategic manoeuvring. However, these works only investigate properties of the dialogue protocols; they do not consider particular strategies for such dialogues as we do here.

7 Concluding Remarks

We have presented a dialogue system for joint deliberation where the agents involved in the decision making may each have different preferences yet all want an agreement to be reached. We defined how arguments and critiques are generated and evaluated, and how this is done within the context of a dialogue. A key aspect concerns how agents' individual reasoning fits within a more global context, without the requirement to completely merge all knowledge. We presented some properties of our system that show when agreement can be guaranteed, and have explored why an agreement may not be reached. Identifying such situations is crucial for conflict resolution and we have discussed how particular steps can be taken to try to reach agreement when this occurs. In future work we intend to give a fuller account of such resolution steps whereby reasoning about other agents' preferences is central.

Ours is the first work to provide a dialogue strategy that allows agents with different preferences to come to an agreement as to how to act. The system allows strategical manoeuvring in order to influence the dialogue outcome, thus laying the important

foundations needed to understand how strategy design affects dialogue outcome when the preferences involved are subjective.

—

References

1. Walton, D.N., Krabbe, E.C.W.: Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. SUNY Press, Albany, NY, USA (1995)
2. McBurney, P., Hitchcock, D., Parsons, S.: The eightfold way of deliberation dialogue. International Journal of Intelligent Systems **22**(1) (2007) 95–132
3. Tang, Y., Parsons, S.: Argumentation-based dialogues for deliberation. In: 4th Int. Joint Conf. on Autonomous Agents and Multi-Agent Systems. (2005) 552–559
4. Dignum, F., Vreeswijk, G.: Towards a testbed for multi-party dialogues. In: AAMAS Int. Workshop on Agent Communication Languages and Conversation Policies. (2003) 63–71
5. Black, E., Atkinson, K.: Dialogues that account for different perspectives in collaborative argumentation. In: 8th Int. Joint Conf. on Autonomous Agents and Multi-Agent Systems. (2009) 867–874
6. Walton, D.N.: Argumentation Schemes for Presumptive Reasoning. Lawrence Erlbaum Associates, Mahwah, NJ, USA (1996)
7. Bench-Capon, T.J.M.: Agreeing to differ: Modelling persuasive dialogue between parties without a consensus about values. Informal Logic **22**(3) (2002) 231–245
8. Atkinson, K., Bench-Capon, T.J.M.: Practical reasoning as presumptive argumentation using action based alternating transition systems. Artificial Intelligence **171**(10–15) (2007) 855–874
9. Wooldridge, M., van der Hoek, W.: On obligations and normative ability: Towards a logical analysis of the social contract. J. of Applied Logic **3** (2005) 396–420
10. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence **77** (1995) 321–357
11. van der Weide, T., Dignum, F., J.-J, Meyer, Prakken, H., Vreeswijk, G.: Personality-based practical reasoning. In: 5th Int. Workshop on Argumentation in Multi-Agent Systems. (2008) 3–18
12. Bench-Capon, T.J.M., Doutre, S., Dunne, P.E.: Audiences in argumentation frameworks. Artificial Intelligence **171**(1) (2007) 42–71
13. Black, E., Hunter, A.: An inquiry dialogue system. Autonomous Agents and Multi-Agent Systems **19**(2) (2009) 173–209
14. Coste-Marquis, S., Devred, C., Konieczny, S., Lagasquie-Schiex, M.C., Marquis, P.: On the merging of Dung's argumentation systems. Artificial Intelligence **171**(10–15) (2007) 730–753
15. Prakken, H.: Coherence and flexibility in dialogue games for argumentation. J. of Logic and Computation **15** (2005) 1009–1040
16. Devereux, J., Reed, C.: Strategic argumentation in rigorous persuasion dialogue. In: 6th Int. Workshop on Argumentation in Multi-Agent Systems. (2009) 37–54
17. Rahwan, I., Larson, K.: Mechanism design for abstract argumentation. In: 5th Int. Joint Conf. on Autonomous Agents and Multi-Agent Systems. (2008) 1031–1038
18. Parsons, S., McBurney, P., Sklar, E., Wooldridge, M.: On the relevance of utterances in formal inter-agent dialogues. In: 6th Int. Joint Conf. on Autonomous Agents and Multi-Agent Systems. (2007) 1002–1009