A Formal Argumentation Framework for Deliberation Dialogues

Eric M. Kok, John-Jules Ch. Meyer, Henry Prakken, and Gerard A. W. Vreeswijk

Department of Information and Computing Sciences,
Utrecht University,
The Netherlands

Abstract. Agents engage in deliberation dialogues to collectively decide on a course of action. To solve conflicts of opinion that arise, they can question claims and supply arguments. Existing models fail to capture the interplay between the provided arguments as well as successively selecting a winner from the proposals. This paper introduces a general framework for agent deliberation dialogues that uses an explicit reply structure to produce coherent dialogues, guides in outcome selection and provide pointers for agent strategies.

Category I.2.11 [**Artificial Intelligence**] Distributed Artificial Intelligence—Languages and structures, multi-agent systems
General Terms Design, Languages
Keywords Argumentation, multi-agent communication, deliberation dialogues

1 Introduction

In multi-agent systems the agents need to work together in order to achieve their personal and mutual goals. Working together means communication and often these dialogues will be on finding consensus over some belief, action or goal. In the last decade frameworks and protocols for such dialogues have been designed using argumentation theory. Walton and Krabbe [14] give a classification of dialogues types based on their initial situation, main goals and participant aims. In a *persuasion dialogue* agents need to find resolution for some conflicting point of view. They will try to persuade the others by forwarding arguments. In *negotiation*, there is not a conflict on some claim, but rather a potential conflict on the division of resources. A deal needs to be made in which each agent tries to get their most preferred resource allocation. *Deliberation* dialogues in contrast, have a significant cooperative aspect. There is a need for action and the agents need to mutually reach a decision. Although agreement is pursued, individual interests also play part.

The literature on argumentation in multi-agent systems has mainly focused on persuasion and negotiation type dialogues. Few systems for deliberation have so far been proposed. The most sophisticated work is that of McBurney et al.

[5] To accommodate deliberating agents, a language and protocol are given that allow for forwarding and discussing of proposals for action. The protocol that they use is liberal in the sense that very few restrictions are imposed on the agents. The modelling of conflicts on beliefs and interests of the agents is limited to the assessment of commitments. It is stated that a voting phase can be used to derive a winner.

It seems that the inquisitive nature of the deliberation process has been well captured in the existing literature. However, the conflicts that arise are left more or less indeterminate. In persuasion, on the other hand, dealing with conflicts is explicitly modelled. Frameworks for these dialogues allow to determine whether given arguments are justified and consequently point out a winner. Such conflicts can be modelled this way for deliberation as well. This can be used to control the deliberation process by maintaining focus on the topic and support the selection of a winning proposal.

For persuasion dialogues, Prakken [10] has proposed a framework that uses an explicit reply structure to capture the relation between arguments. This in turn is used to ensure coherent dialogues as well as to determine the dialogical status of the initial claim. Our framework will be based on this work, adjusting it for use with deliberation dialogues. This will give several advantages. First, proposals can be assigned a status, which can be used to ensure coherent dialogues. Second, the proposed actions can be classified to guide in the selection of a winner. Moreover, the framework will be general to allow for domain specific instantiations and to capture existing protocols in it.

2 The Deliberation Dialogue

A deliberation dialogue commences when the need for action arises. In other words, it needs to be decided upon what action should be taken. A group of people may need to decide where to go for dinner or some automotive company needs to plan what type of car to develop. Agents will need to conceive novel proposals for action and move them in the dialogue. These proposed actions can then be reasoned upon by the agents. If a proposal is unfavourable to the agent it can question it, while it can support the proposal if it seems advantageous. Agents can even express preferences on the proposals. All this is done to influence the dialogue outcome.

In a multi-agent system, deliberation dialogues are only a part of the full communication system. Other types of dialogue, such as argument-based mutual planning [12] or persuasion, can also be part of the system. Deliberation dialogues are thus part of a context. In particular, it commences when in the context the agents belief they mutually need to decide on some action to realize a common goal. Both the goal and need for action can originate from various sources in the context, such as an authority or an earlier dialogue. When the deliberation dialogue starts, agents have, at least in our framework, already agreed on them and can start generating and evaluating proposals.

Agents will have different personal interests and beliefs, because of which conflicts of opinion will come to light during the dialogue. These conflicts can be solved by embedding persuasion-style dialogues. Agents move arguments and question claims to convince other agents. A decision on the winning proposal may be reached through agreement, a voting system or through some authority. Depending on the domain however, both the supplied arguments and the expressed preferences can still be used.

While persuasion is always competitive, deliberation is partially a cooperative process as well. This is expressed in a mutual goal that every agent needs to respect once they accept to engage in deliberation. Support for their proposals needs to show how the action will achieve this common goal. Agents thus need to mediate between their personal opinions and the mutual objective.

As an example, consider a dialogue between three agents that need to find a place for dinner where they will all enjoy the food. They all have an incentive to work towards an agreement on the restaurant, but as the dialogue progresses, differences on beliefs will also need to be resolved.

- a_1: We should go to the local pizzeria.
- a_2: Why should we go there? I propose we go to the nearby bistro.
- a_1: Well, the pizzeria serves tasty pizza's. Why should we go to the bistro?
- a_2: The toppings at the pizzeria are very dull, while the bistro has the best steaks in town.
- a_3: I agree on going to the bistro, because the seafood there is great.
- a_1: The bistro doesn't even server steaks any more.
- a_3: What makes you think the pizza toppings are so dull?
- a_2: Because the menu hasn't been changed for a very long time. We could also just go to pub.
- a_1: No, I don't want to go there.

3 A Formal Deliberation Framework

As explained, our framework will build on the argumentation framework for persuasion dialogues of Prakken, altering and extending it for use with deliberation dialogues. It models persuasion as a dialogue game in which agents make utterances in a communication language while being restricted by a protocol. The utterances, or moves, are targeted at earlier moves. Every reply is either an attacker surrender, forming an explicit dialogue reply structure. The moves contain claims and arguments in the topic language with an argumentation logic. Since it is a framework it allows for various instantiations of the languages and protocol. In the most basic form the protocol is very liberal, only disallowing agents to speak at the same time and requiring that moves are replies to earlier moves. The dialogue terminates when one of the agents cannot make a legal move. The protocol is defined such that there are no legal moves when there is agreement on the original claim.

The explicit reply structure is utilized in two ways. First, moves have a dialectic status. The idea is that a dialogue move is *in* if it is surrendered or else all its attackers are *out*, and that it is *out* if it has an attacker that is *in*. Now the outcome of the persuasion dialogue can be determined based on the dialogical status of the original claim, viz. if at termination this claim is *in* the proponent is the winner. Second, the protocol may be extended with a relevance rule. This compels the agents to stay focussed on the dialogue topic, giving rise to more coherent dialogues.

To make the framework suitable for deliberation dialogues, several modifications are needed. First, multiple agents need to be supported, while the persuasion framework only covers one proponent and one opponent. Several notions, such as relevance, and protocol rules, such as for termination, need to be revised accordingly. Second, there are multiple proposals instead of a single claim to discuss. The communication language needs support for forwarding, rejecting and questioning them. Multiple proposals also means there are multiple dialogical trees to which the agents may contribute. Third, the dialogue outcome is no longer a direct result of the moves. A winning function is needed to select a single action from all actions that are proposed, or possible none if there is no acceptable option.

Now the formal specification for deliberation systems in our framework is introduced. This definition is taken from [10], with the appropriate additions and revisions.

Definition 1 (Deliberation system). A dialogue system for deliberation dialogues is defined by:

- A *topic language* L_t is a logical language closed under classical negation.
- An *argumentation logic* $\mathcal{L}$ as defined in [11]. It is an instance of the Dung [4] argumentation model in which arguments can be formed using inference trees of strict and defeasible rules. Here, an argument will be written as $A \Rightarrow p$ where A is a set of premises and sub-arguments, $\Rightarrow$ is the top inference rule and p is the conclusion of the argument. Such an argument can be attacked by rebutting the conclusion or a sub-argument, by undermining some premise it uses or by undercutting one of the used inference rules.
- A *communication language* L_c, which is a set of locutions $\mathcal{S}$ and two binary relations R_a and R_s of attacking and surrendering reply on $\mathcal{S}$. Every $s \in \mathcal{S}$ is of the form $p(l)$ where p is a performative and $l \in L_t$, $l \subseteq L_t$ or l is an argument in $\mathcal{L}$. R_a and R_s are disjunct and irreflexive. Locutions cannot attack one locution and surrender to another. Finally, every surrendering locution has an *attacking counterpart*, which is an attacking locution in L_c.
- The set $\mathcal{A}$ of agents.
- The set of *moves* M defined as $\mathbb{N} \times \mathcal{A} \times L_c \times \mathbb{N}$ where each element of a move m respectively is denoted by:
  - $\texttt{id}(m)$, the move identifier,
  - $\texttt{player}(m)$, the agent that played the move,
  - $\texttt{content}(m)$, the speech act, or content, of the move,
  - $\texttt{target}(m)$, the move target.

- The set of *dialogues* $M^{\leq\infty}$ is the set of all sequences $m_1, \ldots, m_i, \ldots$ from M, where each i^{th} element in the sequence has identifier i and for each m_i in the sequence it holds if $\texttt{target}(m_i) \neq 0$ then $\texttt{target}(m_i) = j$ for some m_j preceding m_i in d. The set of finite dialogues $M^{<\infty}$ is the set of all those dialogues that are finite, where one such dialogue is denoted by d.
- A *dialogue purpose* to reach a decision on a single *course of action*, which is a $P \in L_t$. P is a proposition stating that some action should be done.
- A deliberation *context* consisting of the *mutual goal* $g_d \in L_t$.
- A *protocol* $\mathcal{P}$ that specifies the legal moves at each point in the dialogue. Formally a protocol on M is a function that works on a non-empty set of *legal finite dialogues* $D \subseteq M^{<\infty}$ and the mutual goal such that $\mathcal{P} : D \times L_t \longrightarrow Pow(M)$. The elements of $\mathcal{P}(d)$ are called the legal moves after d. $\mathcal{P}$ must satisfy the condition that for all legal finite dialogue d and moves m it holds that $d \in D$ and $m \in \mathcal{P}(d)$ iff $d, m \in D$.
- A *turntaking function* $\mathcal{T} : D \longrightarrow \mathcal{A}$ mapping a legal finite deliberation dialogue to a single agent.
- A *deliberation outcome* specified by a function $\mathcal{O} : D \times L_t \longrightarrow L_t$, mapping all legal finite dialogues and the mutual goal g_d to a single course of action α.

This deliberation system specification gives rise to a dialogue game with an explicit reply structure. The types of locutions of L_c that are available to the agents are enumerated in Table 1, each with the appropriate attacking and surrendering replies. The attacking counterpart for each surrendering locution is displayed in the same row. The locutions that deal with proposals (propose, reject, why-propose and prefer) are taken from McBurney et al. while the ones dealing with persuasion (argue, why, retract, concede) are adopted from Prakken's framework. Below the term *proposal move* is used when the $\texttt{content}(m) = \mathit{propose}(P)$, *argue move* is used when the $\texttt{content}(m) = \mathit{argue}(A \Rightarrow p)$, etc.

Argue moves have a well-formed argument in $\mathcal{L}$ as content. If it attacks some other argue move it should defeat the argument contained in that targeted move following the defeat relation of $\mathcal{L}$. All other speech acts have some well-formed formula in L_t as content. Note that for every move m where $\texttt{content} = \mathit{propose}$, *prefer* or *prefer-equal* it holds that $\texttt{target}(m) = 0$ and for all other locutions $\texttt{target} \neq 0$. Specific instantiations of our framework may use a different communication language with different speech acts, as long as the reply relation is defined.

Series of moves that agents make are called turns.

Definition 2 (Turn). A *turn* T in a deliberation dialogue is a maximal sequence of moves $\langle m_i, \ldots, m_j \rangle$ where the same player is to move. A complete deliberation dialogue d can be split up in the sequence of turns $\langle T_1, \ldots, T_k, \ldots, T_n \rangle$ where $k \in \mathbb{N}$ is the turn identifier. A turn thus only has moves from a single player, defined by $\texttt{player}(T)$.

A deliberation dialogue may be represented a set of ordered directed trees.

Table 1. The available speech acts in the communication language L_c

speech act	**attacks**	**surrenders**
propose(P)	*why-propose*(P) *reject*(P)	
reject(P)	*why-reject*(P)	
why-propose(P)	*argue*$(A \Rightarrow p)$	*drop-propose*(P)
why-reject(P)	*argue*$(A \Rightarrow \neg p)$	*drop-reject*(P)
drop-propose(P)		
drop-reject(P)		
prefer(P, Q)		
prefer-equal(P, Q)		
skip		
argue$(A \Rightarrow p)$	*argue*$(B \Rightarrow q)$ where $B \Rightarrow q$ defeats $A \Rightarrow p$ *why*(q) where $q \in A$	*concede*(p) *concede*(q) where $q \in A$
why(p)	*argue*$(A \Rightarrow p)$	*retract*(p)
concede(p)		
retract(p)		

Definition 3 (Proposal tree). For each proposal move m_i in dialogue d a *proposal tree* P is defined as follows:

1. The root of P is m_i.
2. For each move m_j that is a node in P, its children are all moves m_k in d such that $\texttt{target}(m_k) = m_j$.

This is a tree since every move in d has a single target. Now, for any move m in proposal tree P we write $\texttt{proposal}(m) = m_i$.

An example proposal tree is displayed in Fig. 1, which represents a dialogue between three agents. A proposal is moved, questioned and being supported with an argument that in turn had several replies. For each move m_i the number i is its identifier in the dialogue and between brackets the playing agent is noted. Moves in a dotted box are *out*, those in a solid box are *in*.

4 Dialogical Status of a Move

At every point in time, the dialogical status of a move can be evaluated. The use for this is twofold. First, it helps making dialogues coherent through the notion of move relevance. Secondly, the status of proposal moves can later be used during the selection of the final dialogue outcome.

Every move in a proposal tree is always either *in* or *out*. The distinction between attacking and surrendering replies is used here to make the status of moves concrete.

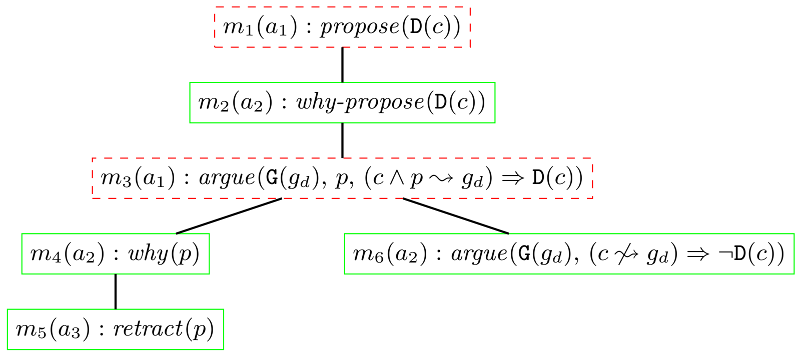

Fig. 1. A small example proposal tree

Definition 4 (Move status). A move m in a dialogue d is *in*, also called *warranted*, iff:

1. m is surrendered in d by every agent $a \in \mathcal{A}$; or else,
2. m has no attacking replies in d that are *in*.

Otherwise it is *out*.

Although this definition is directly taken from [10], special attention here is required to the surrendering attacks. A move is not yet *out* until it is surrendered by every agent in the dialogue, not only by the agent that originally made the attacked move. Take for example the dialogue of Fig. 1. Although agent a_3 moved a *retract*(p) in response to a_2's *why*(p) this targeted move was still *in*. It is not until agent a_1 replied with a *retract*(p) as well that the *why*(p) move is *in* again. A surrendering move is more a statement of no commitment. This idea is made concrete in the following definition of a surrendered move.

Definition 5 (Surrendered move). A move m is *surrendered* in a dialogue d by some agent a iff:

1. m is an argue move $A \Rightarrow p$ and a has made a reply m' to m that has $\mathtt{content}(m') = concede(p)$; or else
2. a has made a surrendering reply to m in d.

Otherwise it is *out*.

The notion of relevance can now be formalised.

Definition 6 (Relevance). An attacking move m in a dialogue d is *relevant* iff it changes the move status of $\mathtt{proposal}(m)$. A surrendering move is relevant iff its attacking counterpart is.

Depending on the domain a different notion of surrendered move or relevance may be useful. Prakken describes a notion of weak relevance that may be adopted. It is weaker in the sense that an agent can contribute multiple ways to change the proposal tree root and still be relevant. This is achieved by only requiring a move to create an additional way to change the status of a proposal. A protocol with weak relevance allows an agent to make multiple attacks per turn in a proposal tree as opposed to a single one if the earlier notion is used, which we below use the term *strong relevance* for.

Definition 7 (Weak relevance). An attacking move m in a dialogue d is *weakly relevant* iff it creates a new or removes an existing *winning part* in the proposal tree P associated with $\texttt{proposal}(m)$ in d. A surrendering move is weakly relevant iff its attacking counterpart is. If the $\texttt{proposal}(m)$ is *in*, a winning part w^P for this tree P is defined as follows:

1. First include the root of P;
2. For each m of even depth, if m is surrendered by every agent $a \in \mathcal{A}$, include all its surrendering replies, otherwise include all its attacking replies;
3. For each m of even depth, include one attacking reply m' that is *in* in d;

The idea of a winning part is that it is 'a reason' why the proposal is *in* at that moment. Since this is not unique, there may be alternative attacking replies, a move is already weakly relevant if it succeeds to create an additional winning part or removes a winning part. Take for example the dialogue of Fig. 1 again. After $argue(\mathtt{G}(g_d), (c \leadsto g_d) \Rightarrow \mathtt{D}(c))$ was moved by agent a_1 there are no more strongly relevant moves in this proposal tree, while there exists new weakly relevant moves, for example $argue(s \Rightarrow g_d)$. This results in a more liberal deliberation process.

5 Turntaking and Termination

We have still not made concrete how agents take turns and when the dialogue terminates.

Definition 8 (Turntaking). Agents take turns in sequence and end their turns explicitly with a skip move. Formally, for a dialogue $d = \langle m_1, \ldots, m_n \rangle$ $\mathcal{T}(d) = \texttt{player}(m_n)$ unless $\texttt{content}(m_n) = skip$ in which case $\mathcal{T}(d) = \texttt{player}(m_n) + 1$.

Clearly, when there are no more legal moves besides the skip move, that is $\mathcal{P}(d) = \{skip\}$, the turn switches. Now, the dialogue terminates if all agents no longer make other moves than directly skipping.

Definition 9 (Termination). A dialogue d terminates on $|\mathcal{A}| + 1$ consecutive skip moves.

The rationale behind the termination rule is that each agent should have the opportunity to make new moves when it stills want to. However, to prevent agents from endlessly skipping until some other agent makes a beneficial move or even a mistake, the number of skip moves is limited.

6 Protocol Rules

Now various protocol rules are discussed. Depending on the domain some might or might not be desirable. First, some rules that prevent agents from playing incoherent moves are added. More precisely, these rules require the agents to be relevant, not to overflow the dialogue.

1. Agents can only reply to moves of others. Formally, for every attacking or surrendering move m in a dialogue $\mathit{player}(m) \neq \mathit{player}(\mathtt{target}(m))$.
2. Every attacking and surrendering move must be relevant.
3. A turn can contain at most one proposal move.
4. A proposal must be unique in the dialogue. Formally, for every proposal move m in d it holds that $\mathtt{content}(m) \notin \{p|p = \mathtt{content}(n)$ of some proposal move $n \in d\}$.

The first rule may be dropped for domains where a more liberal deliberation process is appropriate. This would allow agents to attack their own proposals as well. The relevance of the second rule may be strong or weak relevance. Note that in case of strong relevance there can be at most one attacking move per proposal tree.

Not only the dialogue should be coherent. The same holds for the agents' preference statements on the proposals. A protocol rule is added to ensure that an agent is consistent in his ordering.

5. An agent may only make a prefer move if the resulting option ordering maintains transitivity and antisymmetry. This is further explained below.

The last rules are used to ensure that arguments for (and against) a proposal explain how it (fails to) achieve the mutual goal.

6. Every argue move m with $\mathtt{target}(m) = m'$ and $\mathtt{content}(m') = \mathit{why\text{-}propose}(\mathtt{D}(P))$ will contain an argument in $\mathcal{L}$ with g_d as one of its premises and $\mathtt{D}(P)$ as conclusion.
7. Every argue move m with $\mathtt{target}(m) = m'$ and $\mathtt{content}(m') = \mathit{why\text{-}reject}(\mathtt{D}(P))$ will contain an argument in $\mathcal{L}$ with $\neg g_d$ as one of its premises and $\neg\mathtt{D}(P)$ as conclusion.

The arguments that these protocol rules require are used to make sure that a proposal for action P will indeed (fail to) achieve the mutual goal g_d. Put differently, the proposed action needs to be *appropriate* in relation to our dialogue topic. The topic language and used logic therefore need support to express this. One option, used below, is to include an inference rule for the practical syllogism in our logic $\mathcal{L}$. Similar to [2] a practical reasoning rule will then be used that says 'if g_d is a goal and P will achieve g_d then P is an appropriate proposal for action'. Such arguments, below written as $\mathtt{G}(g_d), P \rightsquigarrow g_d \Rightarrow \mathtt{D}(P)$, can then be moved.

7 Dialogue Outcome

At any moment in time the outcome of the deliberation dialogue can be determined. As the outcome function dictates, this is a single course of action, or no action at all when there is a structural disagreement. To establish this, the options, which are the moved proposals, are first specified and then classified based on their status. This set of proposals is then considered over the agent preferences to determine a winner.

Definition 10 (Options). The dialogue *options* are defined by a function $O : D \longrightarrow Pow(L_t)$ mapping all legal dialogues to a subset of proposals. For any dialogue d the set of options is $O(d) = \{o | o = \texttt{content}(m)$ for each proposal move $m \in d\}$ (below written simply as O). In reverse, $\texttt{move}(o)$ is used to refer to the move in which the option o was proposed.

The proposal moves that introduced the various options have a move status, which will be used to classify the options. Such a classification is any-time and can thus not only be used in selecting the dialogue outcome, but also during the dialogue by agent strategies.

Definition 11 (Option status). An option $o \in O(d)$ for any dialogue d is:

- *justifiable* iff $move(o)$ is *in*,
- *invalid* iff $\texttt{player}(\texttt{move}(o))$ played a move m such that $\texttt{target}(m) = \texttt{move}(o)$ and $\texttt{content}(m) = \textit{drop-propose}(o)$,
- otherwise it is *defensible*.

Justifiable options are proposals that were questioned but were successfully defended. None of the agents was able to build a warranted case against the proposal. Defensible options are proposals that were attacked by some move that is still warranted. These are thus options that might be reasonable alternatives albeit not being properly supported. Invalid options are those that were retracted by the proposing agent. From the perspective of the multi-agent system, the status of each option hints at its acceptability as dialogue outcome. To settle on one of the options they are first ordered according to some preference.

Definition 12 (Option preference). An *option preference* relation $\preceq$ is a partial order of O. This is defined as $o_i \prec o_j$ (strictly preferred) if $o_i \preceq o_j$ but $o_j \not\preceq o_i$ and we have $o_i \approx o_j$ (equally preferred) if $o_i \preceq o_j$ and $o_j \preceq o_i$.

A preliminary ordering on the options can be made. This captures the idea of preferring justifiable options over non-justifiable ones. This may be used during the selection of a dialogue outcome.

Definition 13 (Preliminary ordering). Using the set of all options a partition $O = O_j \cup O_i \cup O_d$ is created such that

- $O_j = \{o | o \in O$ where o is justifiable $\}$,
- $O_d = \{o | o \in O$ where o is defensible $\}$,

– $O_i = \{o | o \in O$ where o is invalid $\}$.

Now $\preceq_p$ is the total *preliminary ordering* over O such that:

– for every two options $o_k, o_l \in O_j$, O_d or O_i it holds that $o_k \approx_p o_l$,
– for every $o_j \in O_j$ and $o_d \in O_d$ it holds that $o_j \prec_p o_d$,
– for every $o_d \in O_d$ and $o_i \in O_i$ it holds that $o_d \prec_p o_i$.

Justifiable proposals are in principle preferred as dialogue outcome over defensible proposals, which in turn are preferred over invalid ones. However, justifiable options should not always be selected as winner over defensible ones. For one, the preferences as moved by the agents using prefer and prefer-equal moves may be taken into account.

Definition 14 (Agent option ordering). Every agent a has a partial *agent option ordering* $\preceq_a$ over O such that for any two options $o_i, o_j \in O$:

– $o_i \prec_a o_j$ if the agent played some move m where $\texttt{content}(m) = \textit{prefer}(o_j, o_i)$,
– $o_i \approx_a o_j$ if the agent played some move m where $\texttt{content}(m) = \textit{prefer-equal}(o_j, o_i)$.

The protocol forces an agent to be consistent in its preference utterances with relation to the strict ordering of options.

When the dialogue terminates, the deliberation dialogue outcome should be selected from the set of options. How this final selection is achieved is totally dependent on the domain and the purpose of the system. For example, there may be an agent authority that gets to choose the winner, an additional phase may be introduced in which agents vote on the outcome or a function may be used to aggregate all (preliminary and agent-specific) preference orderings. In any case we need to leave open the option for *mutual disagreement* [5].

Preference aggregation is extensively studied in the field of social choice theory and is out of the scope of the present paper. [9] It is interesting to note, though, that when maximum social welfare is desirable it may be good to incorporate the notion of our option status in the winner selection. The valuable information obtained during the deliberation dialogue can be used with a public calculus. This would decide on the outcome in a way similar to the use of public semantics and would not need to rely on agents considering these notions in their voting strategies. For single agents, this is already studied in [1]. How to make use of this is left as future research.

8 An Example

To further explain how the different notions work together, consider an example of three agents $\mathcal{A} = \{a_1, a_2, a_3\}$ participating in a deliberation dialogue with mutual goal g_d. We will use all the protocol rules discussed above and adopt a weak form of move relevance. The turns are as follows:

- T_1 by a_1
  m_1 : $propose(\mathtt{D}(z))$ where $z = goToPizzeria$
- T_2 by a_2
  m_2 : $why\text{-}propose(\mathtt{D}(z))$
  m_3 : $propose(\mathtt{D}(b))$ where $b = goToBistro$
- T_3 by a_3
  m_4 : $skip$
- T_4 by a_1
  m_5 : $argue(P \Rightarrow \mathtt{D}(z))$ where
  $P = \{\mathtt{G}(enjoyFood), tastyPizza, goToPizzaria \wedge tastyPizza \rightsquigarrow enjoyFood\}$
  m_6 : $why\text{-}propose(\mathtt{D}(b))$
- T_5 by a_2
  m_7 : $argue(T \Rightarrow \neg\mathtt{D}(z))$ where
  $T = \{\mathtt{G}(enjoyFood), dullTopping, goToPizzaria \wedge dullTopping \rightsquigarrow \neg enjoyFood\}$
  m_8 : $argue(S \Rightarrow \mathtt{D}(b))$ where
  $S = \{\mathtt{G}(enjoyFood), bestSteaks, goToBistro \wedge bestSteaks \rightsquigarrow enjoyFood\}$
- T_6 by a_3
  m_9 : $argue(D \Rightarrow \mathtt{D}(b))$ where
  $D = \{\mathtt{G}(enjoyFood), greatSeafood, goToBistro \wedge greatSeafood \rightsquigarrow enjoyFood\}$
- T_7 by a_1
  m_{10} : $argue(\neg m \Rightarrow \neg s)$ where $m = steakOnMenu$
- T_8 by a_2
  m_{11} : $skip$
- T_9 by a_3
  m_{12} : $why(d)$ where $d = dullTopping$
- T_{10} by a_1
  m_{13} : $skip$
- T_{11} by a_2
  m_{14} : $argue(n \Rightarrow d)$ where $m = menuNeverChanged\}$
  m_{15} : $propose(\mathtt{D}(p))$ where $b = goToPub$
  m_{16} : $prefer(b, p)$ m_{17} : $prefer(p, z)$
- T_{12} by a_3
  m_{18} : $prefer(b, p)$
- T_{13} by a_1
  m_{19} : $reject(p)$
  m_{20} : $prefer(z, b)$
  m_{21} : $prefer\text{-}equal(b, p)$
- T_{14} by a_2
  m_{22} : $skip$
- T_{15} by a_3
  m_{23} : $skip$
- T_{16} by a_1
  m_{24} : $skip$
- T_{17} by a_2
  m_{25} : $skip$

At that point, the proposal trees of the dialogue will look as represented Fig. 2. To see how the dialogical status and protocol rules affected the agents, consider turn T_5, in which agent a_2 tries to refute the proposal for $do(goToPizzeria)$ as made by agent a_1 and support its own proposal for $do(goToBistro)$.

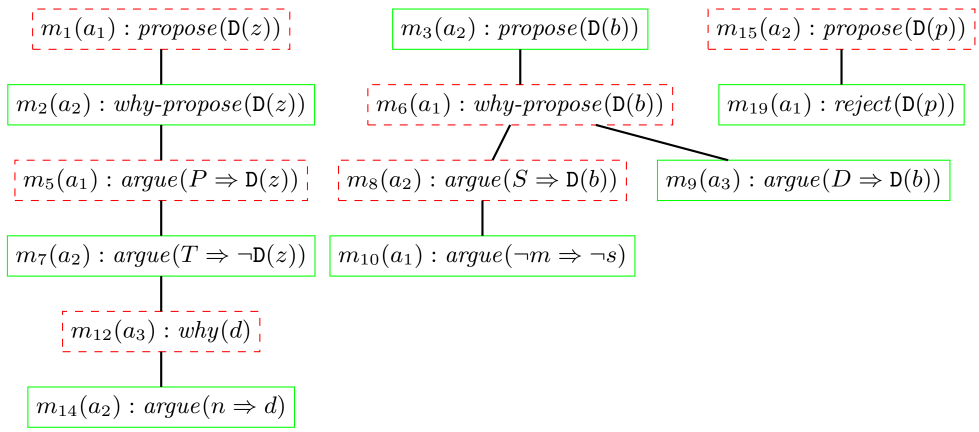

Fig. 2. The proposal trees of the example

To somehow attack proposal D($goToPizzeria$) the agent needs to find a point of attack, which should always be a relevant move. Within this proposal branch, the only points of attack are to attack m_5 or to move another reply to m_1. A relevant move to m_5 can be both an argue (rebuttal, undercutter or underminer) or a why move. Since the proposal move m_1 was already questioned with a *why-propose* the only remaining valid reply there is to move a *reject*(D($goToPizzeria$)). The agent chooses to rebut the conclusion of m_5 with some argument $T \Rightarrow \neg$D($goToPizzeria$).

Within the same turn, the agent also decides to give support to its own proposal D($goToBistro$). To make this proposal *in*, it will have to find a relevant attack move. In this case the only legal attacking move is to forward an argument with conclusion D($goToBistro$) in reply to m_6, which it does in the form $S \Rightarrow$ D($goToBistro$).

Weak relevance is displayed in turn T_6 where agent a_3 make the move *argue*($D \Rightarrow$ D($goToBistro$)). Although at that point a winning part for the proposal tree of D($goToBistro$) already existed, specifically $\{m_3, m_6, m_8\}$, a new winning part $\{m_3, m_6, m_9\}$ is created. If instead strong relevance is used, then move m_9 is not relevant and thus illegal. In turn, the move m_{10} by agent a_1 is only weakly relevant because it removed one winning part in the proposal tree without changing the status of `proposal`(m_8).

The dialogue terminates after turn T_{25}, when agent a_2 was the first to skip twice in a continuous series of skips. The proposal moves of $goToPizzeria$ and $goToPub$ are *out* so those options are defensible. The proposal move of $goToBistro$ on the other hand is *in* and so this option is justifiable. Partitioning the options set O according to the option status results in $O_j = \{goToBistro\}$ and $O_d = \{goToPizzeria, goToPub\}$. This gives a preliminary ordering $goToPizzeria \approx_p goToPub \prec_p goToBistro$. The agent orderings follow directly from the prefer moves they made. The agent option ordering for a_1 is $goToBistro \approx_{a_1} goToPub \prec_{a_1} goToPizzeria$, while that of a_2 is $goToPizzeria \prec_{a_2} goToPub \prec_{a_2} goToBistro$ and the ordering of a_3 is $goToBistro \prec_{a_3} goToPub$.

9 Basic Fairness and Efficiency Requirements

McBurney et al. [6] have proposed a set of 13 desiderata for argumentation protocols. These are criteria which dialogue game protocols need to adhere for basic fairness and efficiency. Each of the desiderata can be verified against our deliberation framework and protocol.

1. **Stated Dialogue Purpose** The protocol is explicitly designed to decide on a course of action.
2. **Diversity of individual purpose** Agents are allowed to have personal goals that possibly conflict with the stated mutual goal.
3. **Inclusiveness** Many agents can join the deliberation dialogue and no roles are enforced upon them.
4. **Transparency** The rules of our framework are fully explained, but it is up to an implementation to make sure every agents knows these rules and knows how to play the game.
5. **Fairness** Every agent has equal rights in the dialogue and the framework allows for fair winner selection methods. Since an agent may always choose not to move (any more) at all, it is never forced to adopt or drop some belief or goal.
6. **Clarity of Argumentation Theory** The reply structure and notion of relevance in our framework are not hidden implicitly in a protocol, but made explicit. Moreover, the structure of arguments is formalised in an explicitly defined argumentation logic and topic language.
7. **Separation of Syntax and Semantics** The communication language is separately defined from the protocol. Also, dialogues in the framework are independent of the agent specification while their public behaviour can still be monitored.
8. **Rule Consistency** We have not studied the rule consistency in detail, but the protocol will never lead to deadlocks; agents can always skip their turn or make a new proposal and within a proposal tree there is always a way to make a new contribution, as long as the top argue move was not conceded.
9. **Encouragement of Resolution** Agents are encouraged to stay focussed on the dialogue topic through the notion of relevance. If agents still have something to say, there is always the opportunity to do so.

10. **Discouragement of Disruption** Disruption is discouraged through the definition of legal speech acts, which are separated in attacks and replies. This restricts the available moves, for example agents cannot attack their own moves. However, it is still possible for aggressive agents to question everything that is claimed and no agent is compelled to accept any claim.
11. **Enablement of Self-Transformation** Agents are allowed to adjust their beliefs or goals depending on the arguments that are moved and preferences that are expressed. Moreover, they are allowed to drop proposals and to retract or concede claims.
12. **System Simplicity** Simplicity of the system is hard to prove or disprove. However, it is highly modular; communication and topic languages are separated and various alternative protocol rules may be adopted or dropped. The winner function is left unspecified, but this may range from a dictator agent to a social welfare-based function.
13. **Computational Simplicity** The computational implications of our framework have not yet been studied. However, the separation of agent and framework designs is at least one step towards simplifying the complexity.

Conforming to these guidelines does not yet mean that every dialogue will be fair and effective. A better understanding is needed of what fair and efficient deliberation dialogues are. Indeed, future work will need to assess how the deliberation process and outcome can be evaluated in relation to the initial situation. In contrast to beliefs, actions will never have an actual truth value but are rather more or less applicable in a specific situation. [5]

New research will also focus on more complete fairness and effectiveness results. For example it is interesting to see how agent attitudes [8] are influential in deliberation dialogues. Moreover, additional formal properties are interesting to study such as the correspondence between the dialogue outcome and the underlying logic of [10].

10 Related Work

The literature on argumentation theory for multi-agent systems includes several attempts at designing systems for deliberation dialogues. Earlier we already briefly discussed the most important work on argumentation in deliberation, i.e. that of McBurney et al. [5] They propose a very liberal protocol for agents to discuss proposals restricted by the advancement of a series of dialogue stages. The used speech acts are very similar to that of our framework, although no explicit logic is used to construct and evaluate arguments. Proposals can be forwarded or rejected, claims and arguments are made, questioned or retracted and preferences are expressed. The resulting commitments of agents are determined, but as in our model they are not used to restrict the legal moves.

Specific support is built into their system for discussion of different perspectives about the problem at hand. Perspectives are influential factors such as moral implications and costs. These perspectives can be integrated in our framework as well though the adopted topic language and logic. One model that could be adopted is proposed in [15].

Agents in the framework of McBureny et al. are constrained in their utterances only by preconditions of the different speech acts. For example, they may not state a preference on two actions before they are asserted. Our model accomplishes this through the explicit reply structure of moves rather than using preconditions. Moreover, our model can enforce dialogical coherence through the notion of move relevance.

To decide on a winning proposal agents need to unanimously accept some proposal or a voting system may be used. This way any knowledge of the arguments on proposals is discarded. In contrast, our model may utilise this knowledge on the multi-agent level to decide on a fair winner without the need for a consensus.

A dialogue protocol on proposals for action is introduced in the work of Atkinson et al. [2] They list all the possible ways to attack a proposal for action, including the circumstances, the goal it achieves and the values it promotes. In our framework, both the goal and action itself are explicitly stated, while the circumstances appear within the arguments that are moved in our deliberation dialogues. As explained earlier, support for values, which are similar to the perspectives of McBurney et al. [5], will be added later.

Many locutions are available to attack proposals, like 'deny goal exists' or 'ask circumstances'. These are needed because no explicit reply structure is present. This also means that no direct relation between the attacks and the resolution of conflicting statements can be made. It is assumed that agents eventually agree on the subject at hand, agree to disagree or use a separate argumentation framework to establish the validity of the proposal. Moreover, the complete work only covers dialogues on a single proposal for action, which makes it persuasion rather than deliberation, albeit being about actions instead of beliefs.

A practical application of multi-agent deliberation dialogues was developed by Tolchinsky et al. [13] A model for discussion on proposals is coupled to a dialogue game. In the model, agents are proponents or opponents of some proposal for action, while a mediator agent determines the legal moves and evaluates moved arguments to see if they are appropriate and how they support or criticize the proposal for action. Although the paper focusses on the translation and application of argument schemes, it is interesting to see how their work can be modelled inside our framework. The number of proposals is limited to a single action, namely to transplant some organ to some recipient, with a mutual goal to find the best organ donor. A dialogue has to start with propose, reject and why-reject moves after which agents can play argue moves. Whether the proposal is also the winner is determined an the authoritative mediator agent.

11 Conclusions

In this paper a framework for multi-agent deliberation dialogues has been proposed. The contribution is twofold.

The general framework for persuasion-type dialogues of Prakken [10] has been altered to provide support for multi-party deliberation dialogues. Consequently, non-trivial modifications have been made to the framework. First, support for moving, criticizing and preferring proposals for action was added. By reusing the explicit reply structure we represent deliberation dialogues as directed multiple trees. Second, the notions of dialogical status and relevance have been adapted for multiple agents. In particular, surrendering replies in a multi-agent context are studied and how strong and weak relevance can still be maintained.

Our framework also improves on the existing work on deliberation dialogues. In contrast with McBurney et al. [5], conflicts of interest are handled through a persuasion-style explicit move status. This allows for varying ways to impose coherence on the deliberating agents. Moreover, the status of proposals is used to define a classification so a preliminary ordering on them can be made. This, together with the agents' explicit preferences, may be used to select a winning proposal.

The framework was checked against the desiderata for multi-agent argumentation protocols. Deliberation systems in our framework will adhere to those basic standards for efficiency and effectiveness. A more rigid study on formal properties of the framework will be valuable here as well as a study on how different agent strategies can affect fairness and effectiveness.

As an extension of our framework, we could allow agents to discuss not only beliefs but also goals, values and preferences. For example, attacking of prefer moves could be allowed, by which a new argument tree is started. A preference-based argumentation framework [7] may be used to in turn evaluate the effect on the dialogical status of proposals. To support discussion on values the topic and communication languages can be extended. One option is to incorporate the work of Black and Atkinson [3], who explicitly allow discussion on promoted values.

Acknowledgments This research was supported by the Netherlands Organisation for Scientific Research (NWO) under project number 612.066.823.

References

1. Amgoud, L., Prade, H.: Using arguments for making and explaining decisions. Artificial Intelligence 173(3-4), 413–436 (2009)
2. Atkinson, K., Bench-Capon, T.J.M., McBurney, P.: A dialogue game protocol for multi-agent argument over proposals for action. Autonomous Agents and Multi-Agent Systems 11(2), 153–171 (2005)

3. Black, E., Atkinson, K.: Dialogues that account for different perspectives in collaborative argumentation. In: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems. pp. 867–874. Budapest, Hungary (2009)
4. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77(2), 321–357 (1995)
5. McBurney, P., Hitchcock, D., Parsons, S.: The eightfold way of deliberation dialogue. International Journal of Intelligent Systems 22(1), 95–132 (2007)
6. McBurney, P., Parsons, S., Wooldridge, M.: Desiderata for agent argumentation protocols. In: Proceedings of the first international joint conference on Autonomous agents and multiagent systems. pp. 402–409. Bologna, Italy (2002)
7. Modgil, S.: Reasoning about preferences in argumentation frameworks. Artificial Intelligence 173(9-10), 901–934 (2009)
8. Parsons, S., Wooldridge, M., Amgoud, L.: Properties and complexity of some formal inter-agent dialogues. Journal of Logic and Computation 13(3), 347–376 (2003)
9. Pini, M.S., Rossi, F., Venable, K.B., Walsh, T.: Aggregating Partially Ordered Preferences. Journal of Logic and Computation 19(3), 475–502 (2008)
10. Prakken, H.: Coherence and Flexibility in Dialogue Games for Argumentation. Journal of Logic and Computation 15(6), 1009–1040 (2005)
11. Prakken, H.: An abstract framework for argumentation with structured arguments. Argument and Computation 1(2) (2010 (to appear))
12. Tang, Y., Parsons, S.: Argumentation-based dialogues for deliberation. In: Proceedings of the 4th International Conference on Multi-Agent Systems. pp. 552–559. ACM (2005)
13. Tolchinsky, P., Atkinson, K., McBurney, P., Modgil, S., Cortés, U.: Agents deliberating over action proposals using the ProCLAIM model. In: Proceedings of the 5th International Central and Eastern European Conference on Multi-Agent Systems. pp. 32–41. Springer-Verlag New York Inc, Leipzig, Germany (2007)
14. Walton, D.N., Krabbe, E.C.W.: Commitment in dialogue: Basic concepts of interpersonal reasoning. State University of New York Press, New York (1995)
15. van der Weide, T.L., Dignum, F.P.M., Meyer, J.J.C., Prakken, H., Vreeswijk, G.A.W.: Practical reasoning using values: Giving meaning to values. In: Proceedings of the 6th International Workshop on Argumentation in Multi-Agent Systems. pp. 225–240. Budapest, Hungary (2009)