Applying Preferences to Dialogue Graphs

Sanjay Modgil [a1] Henry Prakken [b]
[a] *Department of Computer Science, Kings College London*
[b] *Department of Information and Computing Sciences, University of Utrecht and Faculty of Law, University of Groningen*

Abstract. An abstract framework for formalising persuasion dialogues has recently been proposed. The framework provides for a range of speech acts, and protocols of varying levels of flexibility. However, the framework assumes the availability of preference information relevant to determining whether arguments moved in a dialogue *defeat* each other. However, preference information may only become available after the dialogue has terminated. Hence, in this paper, we describe dialogues conducted under the assumption of an *attack* relation that does not account for preferences. We then describe how the resultant dialogue graph can be pruned by a preference relation in order to determine whether the winner of the dialogue is still the winner given the newly available preference information. We also describe a class of protocols that account for subsequent pruning by a preference relation, and show that under a restriction on the pruning, if the player defending the dialogue's main topic is winning the dialogue, then (s)he remains the winner irrespective of the preference relation applied.

1. Introduction

In *persuasion dialogues*, participants attempt to persuade other participants to adopt their point of view. Argumentation based formalisations of such dialogues (reviewed in [3]) generally adopt a game theoretic approach in which speech acts are viewed as moves in a game that are regulated by the game's rules (protocol). A recent formal framework for argumentation based dialogue games [7] abstracts from the specific speech acts, requiring only that they conform to an explicit reply structure; each dialogue move attacks or surrenders some earlier move of the other participant. The current winner of a dialogue can then be determined at any stage, based on the *dialogical status* of moves, the evaluation of which exploits the explicit reply structure. Furthermore, the dialogical status of moves also provides for maintaining the focus of a dialogue when encoding protocols of varying levels of flexibility (*single* and *multi-move*, and *single* and *multi-reply* protocols). The framework's level of abstraction and dialectical nature thus mirrors that of a Dung argumentation framework [4] in which arguments are related by a binary conflict based relation of *attack* or *defeat*. It allows for different underlying logics, and speech acts that can be modelled as nodes in a dialogue graph, whose dialogical status is evaluated based on the reply relations.

[1]This author is supported by the EU 6th Framework project CONTRACT (INFSO-IST-034418). The opinions expressed herein are those of the named authors only and should not be taken as necessarily representative of the opinion of the European Commission or CONTRACT project partners.

For example, consider the initial move of a proponent P claiming: P_1 = "My Mercedes is safe". An opponent O may then query with an attacking reply on P_1: O_2 = "Why is your Mercedes safe". The idea is that O is the current winner, as the burden of proof is on P to reply with an argument for its claim. The *dialogical status* of P_1 is said to be *out*, and O_2 is said to be *in*. P may surrender to O_2 with a reply that retracts the claim P_1, or may in turn attack O_2 with the required argument: P_3 = "My Mercedes is safe since it has an airbag". O_2 is now *out*, P_3 and P_1 *in*, and P is the current winner. O might move a surrendering reply to P_3, conceding the argument's claim, or move an attacking reply to P_3, counter-arguing that Mercedes cars are not safe.

The framework of [7] assumes that if an argument B replies to another argument A, then it is required that B *defeats* A, assuming the standard:

if B attacks A, then B defeats A only if A is not stronger than (preferred to) B.

However, it may be that a preference relation on arguments can only be applied **after** termination of the dialogue. Hence, B may reply to A, but only after the dialogue is it determined that A is in fact stronger than B; information that would have invalidated moving B had it been known at the time of the dialogue. For example in legal adjudication procedures [8] an adjudicator typically decides the dispute after the dialogue between the adversaries has been terminated, so that in that dialogue the preferences that the adjudicator will apply are not yet known. Also, consider the CARREL system [10] developed to demonstrate the *ASPIC* project's argumentation components (www.argumentation.org). CARREL facilitates argumentation between heterogenous agents (software and human) over the validity of organs for transplantation. The dialogues are mediated by a component implementing a protocol in the framework of [7]. Currently, the strengths of arguments are assumed to be given at the time of the dialogue. However, a fully realised CARREL system requires that arguments' strengths be determined after the dialogue. In [11], development of a case based reasoning engine is described. A constructed dialogue graph is input to an algorithm that determines the relative strengths of (and so defeats between) arguments that symmetrically attack. This is based on a comparison with dialogue graphs associated with previous cases, where the same arguments were used, and where the success of the ensuing transplant is then used to weight these arguments in the new graph.

In section 2 of this paper we review the framework of [7] and introduce some new concepts required for section 3. In section 3 we describe dialogues conducted under the assumption that arguments *attack* rather than *defeat* their target arguments. After termination of a dialogue, a preference relation on the arguments moved in a dialogue is then applied to the dialogue graph, removing argue moves (and the moves that follow these) that would have been invalidated had the preference information been available at the time of the dialogue. We then describe protocols that accounts for *attack*s rather than *defeat*s. We then show a result of theoretical and practical interest that holds under certain conditions; viz. if the initial move is *in* in a dialogue, then it remains *in* irrespective of the preference relation applied to the arguments. Finally section 4 concludes with a discussion of future work.

2. A General Framework for Persuasion Dialogues

In the framework of [7], dialogues are assumed to be for two parties; the *proponent* (P) who defends the dialogue topic t and the *opponent* (O) who challenges t.

Table 1. L_c for liberal dialogues

Locutions	Attacks	Surrenders
claim φ	*why* φ	*concede* φ
why φ	*argue* A ($conc(A) = \varphi$)	*retract* φ
argue A	*why* φ ($\varphi \in prem(A)$) *argue* B ($(B, A) \in \mathcal{R}$)	*concede* φ ($\varphi \in prem(A)$ or $\varphi = conc(A)$)
concede φ		
retract φ		

Definition 1 A *dialogue system for argumentation* (dialogue system for short) is a pair $(\mathcal{L}, \mathcal{D})$, where $\mathcal{L}$ is a logic for defeasible argumentation, and $\mathcal{D}$ is a triple $(L_c, \mathbb{P}, C)$ where L_c is a communication language, $\mathbb{P}$ a *protocol* for L_c, and C specifies the effects of locutions in L_c on the participants' *commitments*.

Definition 2 A *logic for defeasible argumentation* $\mathcal{L}$ is a tuple $(L_t, Inf, Args, \mathcal{R})$, where L_t (the *topic language*) is a logical language, Inf is a set of *inference rules* over L_t, $Args$ (the *arguments*) is a set of AND-trees of which the nodes are in L_t and the AND-links are inferences instantiating rules in Inf, and $\mathcal{R}$ is a binary relation on $Args$. For any argument A, $prem(A)$ is the set of leaves of A (its premises) and $conc(A)$ is the root of A (its conclusion).

An argument B *backward extends* an argument A if *conc*(B) = ϕ and $\phi \in$ *prem*(A). The concatenation of A and B (where B *backward extends* A) is denoted by $B \otimes A$.

Defeasible inference in $\mathcal{L}$ is assumed to be defined according to the grounded semantics [4]. Note that the framework abstracts from the nature of the rules in Inf. In this paper, we assume $\mathcal{R}$ denotes either a conflict based *attack* relation, or a *defeat* relation that additionally accounts for preference information. From hereon, we write $A \rightarrow B$ to denote that $(A, B) \in \mathcal{R}$, and $A \leftrightarrow B$ to denote $(A, B), (B, A) \in \mathcal{R}$. In this paper we will ignore commitment rules since they are not relevant to the work presented here. A communication language is a set of locutions and two relations of attacking and surrendering reply defined on this set.

Definition 3 A *communication language* is a tuple L_c = (S,R_a,R_s), where:

S is a set of *locutions* such that each $s \in S$ is of the form $p(c)$, where p is an element of a given set of performatives, and c either is a member or subset of L_t, or is a member of $Args$ (of some given logic $\mathcal{L}$).

R_a and R_s are irreflexive *attacking* and *surrendering reply* relations on S that satisfy:

1. $\forall a, b, c : (a, b) \in R_a \Rightarrow (a, c) \notin R_s$ (a locution cannot be an attack and a surrender at the same time)
2. $\forall a, b, c : (a, b) \in R_s \Rightarrow (c, a) \notin R_a$ (surrenders cannot be attacked since they effectively end a line of dispute)

An example L_c is shown in Table 1 (in which we write '$(A, B) \in \mathcal{R}$' rather than 'A defeats B' as in [7]). For each locution, its surrendering replies and attacking replies are shown in the same row, where the latter are said to be the *attacking counterparts* of the row's surrendering replies. Note however, that for the second line of the *argue* A row, *argue* B is an attacking counterpart of *concede* φ only if the conclusion of B negates or is negated by φ. (So the attacking counterpart of conceding a premise is a premise-attack and the attacking counterpart of conceding a conclusion is a rebuttal.).

In general, the protocol for a communication language L_c is defined in terms of the notion of a dialogue, which in turn is defined with the notion of a move:

Definition 4 Let L_c = (S,R_a,R_s). The set M of *moves* is defined as $\mathbb{N} \times \{P, O\} \times S \times \mathbb{N}$, where the four elements of a move m are denoted by, respectively:

- $id(m)$, the *identifier* of the move,
- $pl(m)$, the *player* of the move,
- $s(m)$, the *locution* performed in the move,
- $t(m)$, the *target* of the move.

The set of (finite) *dialogues*, denoted by ($M^{<\infty}$) $M^{\leq\infty}$, is the set of all (finite) sequences $m_1, \ldots, m_i, \ldots$ from M such that: each i^{th} element in the sequence has identifier i; $t(m_1) = 0$; for all $i > 1$ it holds that $t(m_i) = j$ for some m_j preceding m_i in the sequence.
For any dialogue $d = m_1, \ldots, m_n, \ldots$, the sequence $m_1, \ldots, m_i$ is denoted by d_i, where d_0 denotes the empty dialogue. When d is a dialogue and m a move then d, m denotes the continuation of d with m.

In general we say 'm is in d' if m is a move in the sequence $d = m_1, \ldots, m_i, \ldots$. When $t(m) = id(m')$ we say m *replies to* its target m' in d, and abusing notation we may let $t(m)$ denote a move instead of just its identifier. When $s(m)$ is an attacking (surrendering) reply to $s(m')$ we also say m is an attacking (surrendering) reply to m'.

We now review [7]'s protocol rules that capture a lower bound on the coherence of dialogues. Note that since each move in a dialogue is a reply to a single earlier move, then any dialogue can be represented as a tree. Prior to presenting the protocol rules we define in this paper the notion of a line in a dialogue, and a representation of a finite dialogue as the set of paths (from root node to leaf) or 'lines' (of dialogue) that constitute the dialogue tree whose root node is the initial move m_1.

Definition 5 Let d be a finite dialogue with initial move m_1, and let *leaves*(d) = $\{m|$ m is a move in d, and no m' in d replies to m $\}$.
• Let m_k be any move in d. Then *line*(d, m_k) = $m_1, \ldots, m_k$, where for $j = 2 \ldots k$, $t(m_j) = m_{j-1}$.
• Let *leaves*(d) = $\{m_{k_1}, \ldots, m_{k_n}\}$. Then *lines*($d$) = $\bigcup_{i=k_1}^{k_n}$ *line*(d, m_{k_i}).
• Let $l = m_1, \ldots, m_k \in$ *lines*(d). Then, for $i = 1 \ldots k$, $l' = m_1, \ldots, m_i$ is a sub-line of l.
• If $l = m_1, \ldots, m_i$, $l' = m_1, \ldots, m_i, \ldots, m_j$, then l' is said to be larger than l.

Definition 6 A *protocol* on M is a set $\mathbb{P} \subseteq M^{<\infty}$ such that whenever d is in $\mathbb{P}$, so are all initial sequences that d starts with. A partial function $Pr : M^{<\infty} \longrightarrow \mathcal{P}(M)$ is derived from $\mathbb{P}$ as follows:

- $Pr(d) =$ undefined whenever $d \notin \mathbb{P}$;
- $Pr(d) = \{m \mid d, m \in \mathbb{P}\}$ otherwise.

The elements of *dom*(Pr) (the domain of Pr) are called the *legal finite dialogues*. If d is a legal dialogue and $Pr(d) = \emptyset$, then d is said to be a *terminated* dialogue.
Let T be a turntaking function $T : M^{<\infty} \longrightarrow \mathcal{P}(\{P, O\})$ such that $T(\emptyset) = \{P\}$.
Then ($\mathcal{L}$ = (L_t, *Inf*, *Args*, $\mathcal{R}$), $\mathcal{D}$ = (L_c, $\mathbb{P}$, C)) is a coherent dialogue system if $\mathbb{P}$ satisfies the following basic conditions for all moves m and all legal finite dialogues d.

If $m \in Pr(d)$, then:

- R_1: $pl(m) \in T(d)$;
- R_2: If $d \neq d_0$ and $m \neq m_1$, then $s(m)$ is an attacking or surrendering reply to $s(t(m))$ according to L_c;
- R_3: If m replies to m', then $pl(m) \neq pl(m')$;
- R_4: If there is an m' in d such that $t(m) = t(m')$ then $s(m) \neq s(m')$.
- R_5: For any $m' \in d$ that surrenders to $t(m)$, m is not an attacking counterpart of m' (no move has both a surrender and its attacking counterpart).
- R_6: For any $m' \in d$ such that $pl(m')$ = P, $s(m')$ = $argue\ B$, m' replies to m, $s(m)$ = $argue\ A$, then $argue\ B$ is not in *line*(d, m).

Note that R_2 combined with Table 1's requirement that an *argue* reply must be in the relation $\mathcal{R}$ to its target, effectively builds a version of the argument-game proof theory of [9] into the protocol. Note also that in this paper we have added R_6 to the basic rules of [7] to avoid unnecessary non-termination of dialogues. It is known that this rule (but not the corresponding rule for O) does not change soundness and completeness of the argument game with respect to grounded semantics.

So far any 'verbal struggle' can fit the above framework. It can be specialised for a particular communication language and associated set of protocol rules. Here, we review [7]'s class of *liberal* dialogue systems (parameterised by a logic $\mathcal{L}$) that make use of L_c in Table 1, and in which the participants have much freedom. Two additional protocol rules are added to those in definition 6:

R_7 : If d = $\emptyset$ then $s(m)$ is of the form *claim*(ϕ) or *argue* A (proponent P starts with a unique move that is a claim or argument)

R_8 : if m concedes the conclusion of an argument in m', then m' does not reply to a *why* move (ensuring that only conclusions of counter-arguments can be conceded)

Consider the following dialogue (where ϕ *since* $\alpha_1, \ldots, \alpha_n$ denotes an argument):

Example 1 [Example Dialogue]

P_1: a *since* g, f	
O_2: *concede* f	(O_2 is a surrendering reply to P_1)
O_3: *why* g	(O_3 is an attacking reply to P_1)
P_4: g *since* $\neg b$	(P_4 is an attacking reply to O_3)
O_5: $\neg g$ *since* $\neg g$	(O_5 attack replies to P_4 with the fact $\neg g$)

As discussed in section 1, [7] defines evaluation of the current winner of a dialogue based on the *dialogical status* of moves.

Definition 7 Let a move m in a dialogue d be *surrendered* iff it is an *argue* A move and it has a reply in d that concedes A's conclusion; or else m has a surrendering reply in d. Then, all attacking moves in a finite dialogue d are either *in* or *out* in d. Such a move m is *in* iff

1. m is surrendered in d; or else
2. all attacking replies to m are *out*

Otherwise m is *out*.

We say that if the status of the initial move m_1 in *in* (*out*) then P (O) currently wins d.

In example 1, O_2 is a surrendering reply to P_1, but P_1 is not surrendered since its conclusion has not been conceded. O is the current winner since O_5 is *in*, hence its target P_4 is made *out*, hence O_3 is *in*, and so P_1 is *out*.

During the course of a dialogue d, players implicitly build a dialectical graph g_d = ($Args_d$,$\mathcal{R}_d$) (a Dung graph [4]) of arguments and counter-arguments related to the dialogue topic (see [7] for details of how d_g is constructed). Intuitively, $Args_d$ contains any A in an argue move, backward extended on premises that are not challenged (by *why* locutions) or retracted, provided A does not itself backward extend an argument. $\mathcal{R}_d$ is then defined by the reply relations in the graph. For example, consider the dialogue in figure 1a, and its associated dialectical graph in figure 1b, consisting of two arguments for p constructed during the dialogue.

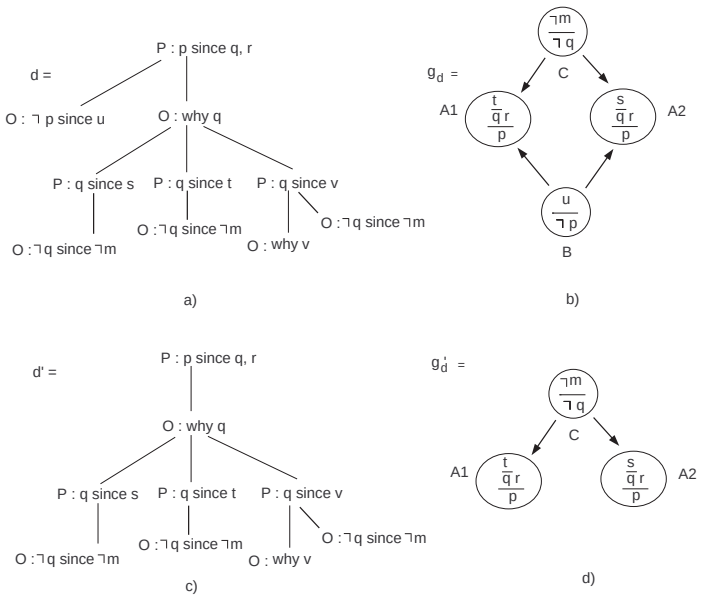

Figure 1. Dialogues and their dialectical graphs. Note that $(p|q, r|v)$ is not in b) and d) since v is under challenge by a *why* locution.

3. Applying Preferences to Dialogues

3.1. Preference based Resolution of Dialogue Graphs

Thus far we have assumed that if a move *argue* B replies to *argue* A, then $(B, A) \in \mathcal{R}$ in the underlying logic, where $\mathcal{R}$ denotes either an *attack* or *defeat* relation. As discussed in section 1, scenarios where a preference relation on arguments (based on their relative strengths) is available only after the dialogue, require a focus on dialogues where $\mathcal{R}$ denotes *attack*. As mentioned earlier, the framework of [7] assumes that backwards extending of arguments does not weaken arguments. This assumption is made explicit here.

Definition 8 Let d be a dialogue in the dialogue system ($\mathcal{L} = (L_t, Inf, Args, \mathcal{R}), \mathcal{D}$). We say that a preordering $Pref$ on $Args$ is a *preference relation* on $Args$, where:
$\gg^{Pref}$ is the associated strict ordering ($A \gg^{Pref} B$ iff $APref\,B$ and not $BPref\,A$)
$\equiv^{Pref}$ is the associated relation of equivalence ($A \equiv^{Pref} B$ iff $APref\,B$ and $BPref\,A$)
We say that $Pref$ is *not weakened by backward extending* iff:
$\forall A, B \in Args$, if $A \gg^{Pref} B$, then $\forall A', B'$ s.t. $A' \otimes A, B' \otimes B \in Args$, $A' \otimes A \gg^{Pref} B' \otimes B$

From hereon we will, unless stated otherwise, assume that $Pref$ is *not weakened by backward extending*. Notice that the backward extension restriction is satisfied by the last link valuation of argument strength (used in [9] and the CARREL [10] system described in section 1). In figure 1a), suppose (*p since q,r*) $\gg^{Pref}$ (¬*p since u*). If $Pref$ is not weakened by backward extending, then $A1 >^{Pref} B$ and $A2 >^{Pref} B$ in the dialogue's dialectical graph.

We now define a resolution *resolve*$(d, pn, Pref)$ of a dialogue d based on $Pref$ and a pruning function pn, where pn takes as input a dialogue line l and a strict ordering $\gg$, and returns a sub-line of l. One can then determine the status of the initial move m_1 in the resolution *resolve*$(d, pn, Pref)$, and thus determine whether, based on the available preference information, proponent remains as the winner or loser of the dialogue.

Definition 9 Let d be a dialogue and *lines*(d) = $\{l_1, \ldots, l_n\}$. Then:
$\bigcup_{i=1}^{n} pn(l_i, \gg^{Pref})$ is the resolution of d w.r.t. $Pref$, denoted *resolve*$(d, pn, Pref)$.

We now define a specific pruning function *prune*. Firstly, note that some argument game proof theories proposed for the grounded semantics (e.g. [1,9]) account for the extra burden of proof on P, by requiring that P's arguments *strictly* defeat their targets (this requirement is not needed for soundness and completeness but makes dialogues shorter and therefore protocols more efficient). Recall that B defeats A if it attacks A, and A is not strictly preferred to B, and B *strictly* defeats A if B defeats A and A does not defeat B. Hence, if A and B symmetrically attack, then B *strictly* defeats A only if $B \gg^{Pref} A$. Hence, the pruning function will only retain a B moved by P as a reply to the symmetrically attacking A, if $B \gg^{Pref} A$.

Definition 10 Let d be a dialogue in ($\mathcal{L} = (L_t, Inf, Args, \mathcal{R}), \mathcal{D}$). Let $l \in$*lines*(d) and $\gg$ a strict ordering on the arguments moved in l. Then l' = *prune*$(l, \gg)$ is the largest sub-line of l such that for all m' in l', if m' replies to m, and $s(m)$ = *argue* A, $s(m')$ = *argue* B, then:

1. it is not the case that $A \gg B$ (it is not the case that B does not defeat A according to $\gg$)
2. if $pl(m')$ = P, $pl(m)$ = O, and $(A, B) \in \mathcal{R}$, then $B \gg A$ (B strictly defeats A according to $\gg$)

Example 2 Consider dialogue d in figure 1a). The initial *argue* move *p since q,r* is *out*. Suppose (*p since q,r*) $\gg^{Pref}$ (¬*p since u*), and (¬*q since* ¬ *m*) $\equiv^{Pref}$ (*q since s*).
d' = *resolve*(*d*,*prune*,*Pref*) is shown in figure 1c) in which *argue* move *p since q,r* remains *out*. d''s dialectical graph g'_d is shown in figure 1d). Note that since $Pref$ is not weakened by backward extending, then $A1 \gg^{Pref} B$ and $A2 \gg^{Pref} B$, suggesting that g'_d might be directly obtained by applying $Pref$ to g_d (although this is not trivial and remains a topic for future work).

Finally, note that in [7] it is shown that for dialogues without surrenders in which the players play logically perfectly (the players move any attacking arguments defined by the dialogue's dialectical graph, as for example shown in figure 1a)), the initial move is *in* iff the main argument is in the grounded extension of the argument's dialectical graph. It immediately follows from the proofs in [7] that this result still holds for resolved dialogue graphs that satisfy the same properties.

3.2. 'Rational' Protocols

Consider a liberal dialogue d where A, B and C are the arguments moved, and assume $C \leftrightarrow B \rightarrow A$. Suppose the dialogue starts with:

$$d_2 = P_1 : \textit{argue } A, O_2 : \textit{argue } B.$$

As previously observed, argument games for the grounded semantics would prohibit P_3 : *argue* C (given that $B \rightarrow C$). As required, P_1 is *out* since A is not in the grounded extension. However, assuming $\rightarrow$ denotes an attack relation, and subsequent application of preferences, then a 'rational' proponent would move P_3 : *argue* C, so that if it turns out that $C \gg^{Pref} B$, then P_1 is now *in* in the dialogue:

$$d_3 = P_1 : \textit{argue } A, O_2 : \textit{argue } B, P_3 : \textit{argue } C.$$

However, note now that proponent is the current winner of the dialogue, despite the fact that A is not in the grounded extension of $C \leftrightarrow B \rightarrow A$. A rational opponent would move B again. We therefore state that an opponent plays a dialogue *rationally* if the dialogue satisfies the following:

Definition 11 [Rational Opponent] R_{RO}: If d contains a subsequence of moves $\ldots, m_j, \ldots, m_k, \ldots$ such that m_k replies to m_j, and $m_j = O$: *argue* A, $m_k = P$: *argue* B, then: if $(A, B) \in \mathcal{R}$, and there is no m_l in d that replies to m_k such that $m_l = O$: *argue* A, then $m = O$: *argue* A replies to m_k.

Hence, *argue* A is made *out* in:

$$d_4 = P_1 : \textit{argue } A, O_2 : \textit{argue } B, P_3 : \textit{argue } C, O_4 : \textit{argue } B,$$

Suppose d_4 continues with proponent's attacking reply querying a premise ϕ in the second instance of B, obtaining:

$$d_5 = P_1 : \textit{argue } A, O_2 : \textit{argue } B, P_3 : \textit{argue } C, O_4 : \text{argue } B, P_5 : \textit{why } \phi.$$

Observe now that if $B \gg^{Pref} C$, then *resolve*(d_5, *prune*, $Pref$) = P_1 : *argue* A, O_2 : *argue* B in which A is *out*! Proponent's attacking reply on B has been 'lost' in the pruning. A rational proponent would therefore challenge ϕ in the first instance of B. We therefore also state that a proponent plays a dialogue *rationally* if it always replies to the first occurrence in a dialogue line of an argue move by opponent. It is easy to see that this will never deny proponent any opportunity, since any reply that is effective on a later occurrence will be effective on the first occurrence.

Definition 12 [Rational Proponent] R_{RP}: If m_k is a move O : *argue* A in a dialogue d, and m replies to m_k, then letting *line*$(d, m_k) = m_1, \ldots, m_k$, for $i < k$, $m_i \neq O$: *argue* A.

Liberal dialogues in which opponent and proponent play rationally are from hereon referred to as *rational liberal dialogues*.

Intuitively, we would want that the resolution d' of a rational liberal dialogue d by $Pref$, is itself a dialogue that would have been obtained assuming availability of $Pref$ and a defined defeat relation. However, the requirement that arguments moved by P strictly defeat their target arguments is not captured by protocol rules $R_1 - R_8$. We therefore refer to *grounded liberal* protocols that are additionally defined by the rule:

R_9 : If m = P : *argue* B replies to m' = O : *argue* A, then $(A, B) \notin \mathcal{R}$, $(B, A) \in \mathcal{R}$.

In the following proposition we say that line l is pruned at m_j to obtain l' ending in m_i, if $l = m_1, \ldots, m_i, m_j, \ldots$ and $l' = m_1, \ldots, m_i$.

Proposition 1 Let d be a rational liberal dialogue in ($\mathcal{L} = (L_t, Inf, Args, \mathcal{R}), \mathcal{D} = (L_c, \mathbb{P}, C)$) and let d' = *resolve*(d,*prune*,*Pref*).
Let $Pref$ be a preordering on $Args$ and $\mathcal{R}'$ = $\{ (B, A) \mid (B, A) \in \mathcal{R}$ and it is not the case that $A \gg^{Pref} B \}$. Then:
d' is a dialogue in ($\mathcal{L}' = (L_t, Inf, Args, \mathcal{R}'), \mathcal{D}' = (L_c, \mathbb{P}', C)$), where $\mathbb{P}'$ is a grounded liberal protocol, and:
for every $l \in$ *lines*(d) pruned at m_j to obtain l' ending in m_i, m_j is not a legal reply to m_i in $\mathcal{D}'$.

Proof: Let $l = m_1, \ldots, p$: *argue* A, $\overline{p}$: *argue* B,... be any line in *lines*(d), and suppose l is pruned to obtain $l' = m_1, \ldots, p$: *argue* A. There are two cases to consider:
1) $A \gg B$, and i) $(B, A) \in \mathcal{R}$, $(A, B) \notin \mathcal{R}$, or: ii) (B, A),$(A, B) \in \mathcal{R}$, and $p = P$, $\overline{p} = O$. Hence, $(B, A) \notin \mathcal{R}'$ so that *argue* B is not a legal reply to *argue* A in the grounded liberal protocol.
2) $p = O$, $\overline{p} = P$, (B, A),$(A, B) \in \mathcal{R}$, and it is not the case that $B \gg A$. In which case $(A, B) \in \mathcal{R}'$.
In both cases (by definition of L_c (table 1) in the case of 1) and R_9 in the case of 2)) $\overline{p}$ cannot move *argue* B as a reply to p : *argue* A in the grounded liberal game.

3.3. Applying Preferences Exclusively to Symmetric Attacks

We now go on to show a practically useful result that holds for rational liberal protocols. The result holds for argumentation formalisms in which preferences are applied only to symmetric attacks. Hence we give the following definition of *pruneS* (*S* for symmetric) that can be used in defining the resolution of a dialogue as described in definition 9.

Definition 13 Let d be a dialogue in ($\mathcal{L} = (L_t, Inf, Args, \mathcal{R}), \mathcal{D}$), $l \in$*lines*(d) and $\gg$ a strict ordering on the arguments moved in l. Then l' = *pruneS*($l, \gg$) is the largest sub-line of l such that for all m' in l', if m' replies to m, and $s(m)$ = *argue* A, $s(m')$ = *argue* B, **and** (A, B)**,** $(B, A) \in \mathcal{R}$, then:

1. it is not the case that $A \gg B$
2. if $p(m')$ = P, $p(m)$ = O, and $(A, B) \in \mathcal{R}$, then $B \gg A$

Proposition 2 below states that under the assumption of rational liberal protocols, and application of preferences to symmetric attacks, then if the status of the initial move m_1 is *in*, then it is *in* **irrespective** of the preference ordering applied. Prior to showing the proposition we define the *line status* of a move and establish a lemma (notice that surrendering replies only terminate lines since they cannot be replied to).

Definition 14 Let $l = m_1, \ldots, m_n$ be a line in a finite dialogue d ($l \in$ *lines*(d)).
- If m_n is a surrendering reply to m_{n-1}, then *line-status*(l,m_{n-1}) is *in*, and for $i < n-1$: *line-status*(l,m_i) is *in* if *line-status*(l,m_{i+1}) is *out*, else *line-status*(l,m_i) is *out*.
- If m_n is an attacking reply to m_{n-1}, then *line-status*(l,m_n) is *in*, and for $i < n$: *line-status*(l,m_i) is *in* if *line-status*(l,m_{i+1}) is *out*, else *line-status*(l,m_i) is *out*.

Lemma 1 Let the dialogical status of the initial move m_1 in dialogue d be *in*. Let *lines*(d) = $\{l_1, \ldots, l_n\}$, and *resolve*(d, pn, $Pref$) = $\{l'_1, \ldots, l'_n\}$. Then, if for $i = 1, \ldots, n$, if for all m in l'_i such that m is in l, either:

- *line-status*(l'_i,m) = *line-status*(l_i,m); or
- if *line-status*(l'_i,m) $\neq$ *line-status*(l_i,m), then either
  - *line-status*(l'_i,m) = *in*, and $p(m) = P$, or
  - *line-status*(l'_i,m) = *out* and $p(m) = O$

then the dialogical status of the initial move m_1 in the dialogue d' corresponding to $\{l'_1, \ldots, l'_n\}$ is *in*.

Proof: Obvious.

Proposition 2 Let $d = m_1, \ldots, m_n$ be a rational liberal finite dialogue in ($\mathcal{L} = (L_t, Inf, Args, \mathcal{R}), \mathcal{D}$). If the dialogical status of m_1 is *in*, then for all preorderings $Pref$ on $Args$, m_1 is *in* in *resolve*(d, *pruneS*, $Pref$).

Proof: Suppose some line $l \in$ *lines*(d) and arguments A, B such that (A, B), $(B, A) \in \mathcal{R}$ and either $A \gg^{Pref} B$, or it is not the case that $B \gg^{Pref} A$. There are three cases where l is pruned to obtain l' (we simply write the players and the arguments):

1. $m_1, \ldots, P_i : A, O_{i+1} : B, \ldots$
2. $m_1, \ldots, O_i : A, P_{i+1} : B, O_{i+2} : A$ (by R_{RO})
3. $m_1, \ldots, O_i : B, P_{i+1} : A, O_{i+2} : B$ (by R_{RO})

(Note that $\ldots, P_i : B, O_{i+1} : A, P_{i+2} : B \ldots$, is excluded by R_6).
i) Suppose $A \gg^{Pref} B$.
In case 1), l' terminates in $P_i : A$, and so the line status of $P_i : A$ is *in*. Hence the line status of every move by P in l' is *in*, every move by O is *out*.
In case 2) the line-status of $O_{i+2} : A$ is *in* since by R_{RP} proponent does not reply to O_{i+2}. Hence, the line-status of moves $P_{i+1} : B$ and $O_i : A$ are *out* and *in* respectively. The line-status of $O_i : A$ in l' ending in $O_i : A$ remains *in*. Hence, the line status of moves in l' are the same as in l.
In case 3) the line-status of $O_{i+2} : B$ is *in* given R_{RP}. Hence, $P_{i+1} : A$ is *out*, $O_i : B$ is *in*. The line status of $P_{i+1} : A$ in l' ending in $P_{i+1} : A$, and so every move by P in l', now changes to *in*, and $O_i : B$, and so every move by O in l', changes to *out*.
ii) Suppose it is not the case that $B \gg^{Pref} A$.
In case 2), l is pruned to l' ending in $O_i : A$, and the line status of $O_i : A$ in l and l' is *in*. Hence, the line status of moves in l' are the same as in l.
Given **i)** - **ii)**, then by lemma 1 the dialogical status of m_1 is *in* in *resolve*($d, Pref$)

The above result is of practical as well as theoretical interest. Preferences only need to be applied to decide the issue when the current winner is the opponent. Furthermore, the restriction on application of preferences to symmetric attacks is satisfied by a number of formalisms. For example, logic programming formalims such as [9] and [5], that

underpin the CARREL system [10]. In these formalisms, A asymmetrically attacks B, only if A claims (proves) what was assumed non provable (through negation as failure) by a premise in B. A then defeats B irrespective of whether B is preferred to A.

Example 3 For the dialogue in figure 1a), suppose that in the underlying logic any two arguments with logically contradictory conclusions symmetrically attack. Suppose O's argument C for $\neg q$ where $\neg$ in $\neg m$ denotes negation as failure ('m is not provable'). Now suppose:
- P replies to each instance of C with the move *argue* m *since* f, where m *since* f asymmetrically attacks C in the underlying logic.
- P replies to $\neg p$ *since* u with *why* u.
p *since* q, r is now *in* and remains *in* irrespective of the preferences between the symmetrically attacking pairs of arguments for p and $\neg p$, and q and $\neg q$

3.4. Non-repetition with symmetric attacks

Finally, a further result can be proven if all attacks between arguments are symmetric, whether dialogues are rational or not. Consider in addition to R_1 - R_8, rule R_{10} defining liberal 'non-repetition' dialogues:

> R_{10}: if d contains a line with O_i : *argue* A; P_{i+1} : *argue* B then m does not reply to P_{i+1} with *argue* A.

This says that O may not move A in reply to argument B if B in turn is a reply to A (for P this is already excluded by R_6). It can now be shown that after resolution the result will be same with our without this new protocol rule.

Proposition 3 Let d be a liberal dialogue where $(A, B) \in \mathcal{R}$ implies $(B, A) \in \mathcal{R}$ in the underlying logic. Let d' be a liberal non-repetition dialogue, obtained from d by pruning every line at a move that violates R_{10}. Then after pruning both dialogues have the same winner.
Proof: We only need to consider cases where a line l in d contains $m_1, \ldots, O_i : A; P_{i+1} : B$. Irrespective of whether we have $O_{i+2} : A$ or not, if $A \gg^{Pref} B$ then both prunings yield lines ending in $O_i : A$, if $B \gg^{Pref} A$ then both prunings yield lines ending in $P_{i+1} : B$, and if it is not the case that $B \gg^{Pref} A$, then both prunings yield lines ending in $O_i : A$.

Note that the above result does not hold if we generalise R_{10} to preclude repetition of argue moves by O in the same line (i.e., as R_6 does for P). Finally, note that non-repetition protocols not only ensure shorter dialogues, but are also more 'realistic' in the sense that it is somewhat counter-intuitive for a player to repeat an argue move that it has already submitted.

4. Conclusions

In this paper we have described a procedure for applying preferences to dialogues in [7]'s general framework. Preference information unavailable at the time of a dialogue can then be used to determine the winner of the dialogue after termination. We described protocols that account for arguments attacking rather than defeating their targets, and where the subsequent resolution of a dialogue graph based on preference information yields a

result that would have been obtained by the defined defeat relation. We also showed that if preferences are applied only to symmetric attacks, then if proponent is currently winning the dialogue, then he wins irrespective of the preference relation applied. Requirements for applying preferences to dialogue graphs were highlighted by implementations of protocols instantiating the framework that have been deployed in the CARREL system [10], and is also relevant for models of legal adjudication procedures [8]. Future work aims to extend the former implementations to enable application of preferences in the manner described in this paper.

The framework of [7] assumes that backward extension does not weaken arguments, thus presupposing the 'last link' evaluation of argument strength [9]. This might suggest that preferences cannot be applied to dialogues formalised in the 'Toulouse-Liverpool' approach [2,6], where the 'weakest link' valuation of arguments violates this principle. However, this approach does not effectively allow for backward extending of arguments. Participants make claims, and when queried are required to defend these claims with Dung acceptable arguments constructed from their shared commitments and individual belief bases. Thus any such argument moved will already be backward extended to the extent that it can be. Any premise challenged will elicit an alternative argument for that premise. However, application of preferences to such dialogues will require extending the pruning mechanisms described in this paper. This is because the dialogues in [2,6] do not explicitly model argue moves replying to argue moves; rather, this reply relation is implicit, in that claims or assertions of propositions reply to each other, where the arguments for these assertions are elicited by why moves.

Acknowledgements This work was partly funded by the EU 6th framework projects ASPIC (FP6-002307) and CONTRACT (FP6-034418).

References

[1] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

[2] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proc. 4th International Conference on MultiAgent Systems (ICMAS-00)*,31–38, Boston, MA, 2000.

[3] ASPIC. Deliverable d2.1: Theoretical frameworks for argumentation. http://www.argumentation.org/Public_Deliverables.htm, June 2004.

[4] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[5] S. Modgil, P. Tolchinsky, and U. Cortés. Towards formalising agent argumentation over the viability of human organs for transplantation. In *4th Mexican Int. Conf. on Artificial Intelligence*, 928–938, 2005.

[6] S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal interagent dialogues. In *Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems (AAMAS-02)*, 394–401, 2002.

[7] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.

[8] H. Prakken. A formal model of adjudication. In S. Rahman, editor, *To appear in: Argumentation, Logic and Law*. Springer Verlag, Dordrecht, 2007.

[9] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.

[10] P. Tolchinsky, U. Cortés, S. Modgil, F. Caballero, and A. Lopez-Navidad. Increasing Human-Organ Transplant Availability: Argumentation-Based Agent Deliberation. *IEEE Special Issue on Intelligent Agents in Healthcare*, 21(6):30–47, 2006.

[11] P. Tolchinsky, S. Modgil, U. Cortés, and M. Sánchez-Marré. CBR and argument schemes for collaborative decision making. In *Proc. 1st International Conference on Computational Models of Argument*, 71–82, 2006.